Advances in Mathematics Education

Egan J. Chernoff
Bharath Sriraman *Editors*

Probabilistic Thinking

Presenting Plural Perspectives

Springer

Advances in Mathematics Education

For further volumes:
www.springer.com/series/8392

Egan J. Chernoff • Bharath Sriraman
Editors

Probabilistic Thinking

Presenting Plural Perspectives

Springer

Editors
Egan J. Chernoff
College of Education
University of Saskatchewan
Saskatoon, Saskatchewan, Canada

Bharath Sriraman
Dept. of Mathematical Sciences
The University of Montana
Montana, MT, USA

ISSN 1869-4918
ISSN 1869-4926 (electronic)
Advances in Mathematics Education
ISBN 978-94-007-7154-3
ISBN 978-94-007-7155-0 (eBook)
DOI 10.1007/978-94-007-7155-0
Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2013956368

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Series Preface

The series *Advances in Mathematics Education* aims mainly to produce monographs based on important issues from ZDM – The International Journal on Mathematics Education from the past; however, "the series is open to proposals from the community on other topics of interest to the field." Probability has, more recently, become a mainstream strand within worldwide mathematics curriculum and, further, is a continual burgeoning area of research in mathematics education. Given the former and latter points, the seventh volume in the series *Advances in Mathematics Education* deals with probabilistic thinking.

This seventh volume exemplifies that *Advances in Mathematics Education* accepts proposals from the community on topics of interest to the field while preserving many of the characteristics of ZDM (since its inception in 1969). As mentioned in the Series Preface to the first book in the series on *Theories of Mathematics Education*, the publication of themed issues as characteristic of ZDM aims to bring the state-of-the-art on central sub-domains within mathematics education. Similarly, this volume is thematic; in that it is based on four different, yet interrelated "perspectives" or central sub-domains within probability: Mathematics and Philosophy, Psychology, Stochastics, and Mathematics Education. Further, this volume continues "the usage of the ancient scholarly Chinese and Indian traditions of commentaries" (Kaiser and Sriraman 2010, p. vi) and solicits "commentaries from experts and novices (ibid.)." As is also the case in past volumes, "prefaces to chapters set the stage for the motivation, purpose and background of a given [perspective]" (ibid.). Lastly, although not directly based on previously published themed issues of ZDM, this volume relies on the work of many authors who were also involved in a recently themed issue of ZDM, "Probability and Reasoning about Data and Risk," which was guest edited by Rolf Biehler and Dave Pratt. Ultimately, this volume demonstrates that the first book in this series *Theories of Mathematics Education* did provide a prototype of the books series, which can be applied to monographs based on important ZDM issues of the past and proposals from the community on other topics of interest to the field.

In what has turned out to be an implicit feature of the *Advances in Mathematics Education* series, the scope of this volume is substantial: 28 chapters, from (in total)

56 authors (including luminaries from the fields of mathematics, psychology and mathematics education) from across the globe. While research has strayed more towards the realm of statistical thinking and reasoning, this book deliberately takes the path less chosen, that is, explores the roots and different facets of probabilistic thinking and, in the forward looking spirit of the series, fertile directions in which the research can be pushed. *Probabilistic Thinking* is both anthological and future-oriented, with the explicit purpose of becoming a reference book for mathematics education.

Hamburg, Germany Gabriele Kaiser
Missoula, USA Bharath Sriraman

To
Kristen and Scout for their unconditional support
&
Sarah, Jacob and Miriam—for letting me imagine probabilities against all odds

The Most Common Misconception About Probability?

DISCLAIMER: I have not carried out a systematic study to catalog misconceptions about probability, so I can't claim to know if there is a single most common one, and if so what it is. But I can offer a likely candidate, and it has significant implications for the way we teach probability. I would put a high probability on being able to delete the question mark from my title and produce an accurate description.

Misunderstandings about probability are legion. It is, after all, a fiendishly complex and elusive notion to wrap your mind around. Yet as the only reliable means we have to predict—and plan for—the future, it plays a huge role in our lives, so we cannot ignore it, and we must teach it to all future citizens. The purpose of this article is to highlight one important goal of that teaching.

Why only one? Because it addresses a widespread confusion at a very fundamental level, and it is, I think, a confusion we can avoid if we are careful how we introduce the notion of probability.

Since the early 1980s, I have written regularly on mathematical topics for various newspapers, magazines, and in more recent years various online publications and blogs, and in the process have received a great deal of correspondence. Whenever I write about probability, a small flood of correspondence generally ensues.

Some of the feedback I get is predictable, both in amount and the nature of the comments. Famous probability puzzles such as the "What is the probability that my other child is a girl?" question and the Monty Hall teaser always generate a lot of controversy, as a new group of innocents encounter them for the first time. But puzzles such as those are designed to generate such a response, and depend on a carefully contrived set of background assumptions (only some of which are usually articulated) in order to guarantee the counter-intuitive "correct" answer the questioner has in mind. Giving the "wrong" answer to such a puzzle does not necessarily mean the respondent does not understand probability theory. In fact, a person who does have such understanding can invariably provide a sound rationale in support of their answer by articulating a background condition the questioner did not explicitly exclude.

For example, in the child gender problem, I say that "I have two children and one of them is a girl, what is the probability that my other child is a girl?" The question is phrased so that the listener will likely answer, incorrectly, "One half." The "correct" answer is supposed to be "One third," arrived at as follows. Since children are born serially, there are four possibilities, ordered by birth, BB, BG, GB, GG. (Many people make a type-token error at this point and say I should list BB and GG twice each, but that is a separate issue.) My statement eliminates the BB possibility, leaving BG, GB, GG. Of these three, only GG results in my other child being a girl, giving the probability $1/3$.

That, at least, is how this example is usually presented and solved. But it is possible to argue for other answers. For example, what if I am a compulsively logical person who always provides information about my life in chronological order, and my listener knows that about me? Then my statement implies that my first child is a girl, so the possibilities are GB and GG, and then the answer is $1/2$.

Alternatively, suppose my listener and I come from a strictly patriarchal society where parents always give priority to information about sons. Then my statement implies that I have no boys, and the answer is 1.

True, these two contextual circumstances are contrived, but then so is the original puzzle. In real life, the accepted norms and maxims of communication generally prohibit making a statement of the form "At least one of my children is..." Since probability theory is an eminently practical subject, intentionally used widely in real-world situations, confusion over contrived puzzles does not mean someone cannot make reliable practical use of probability.

Leaving aside such mind-benders, however, over the years, exchanges with readers have shown me that there is a common misconception about the very nature of probability. It is this. Many people believe that *events*—that is, things that happen in the world—have a unique probability.

For example, in exchanges about the child-gender puzzle, where I have discussed the effects of different background assumptions, readers often would say, "Yes, I see that, but what is the *actual* probability"? To which, of course, I say that the actual probability is 1. I actually do have two children, both girls. Their gender is not a probabilistic matter. They were born long ago, and their genders factually established at birth, both in reality and in law.

The point is, I go on to explain, probabilities apply not to events in the real world we are familiar with, but to *our information* about that world at any given moment in time.[1] This crucial aspect of the notion of probability is often ignored, if not obscured, by the way we normally introduce the concept to students, using easily understood, and readily implemented, experimental procedures such as tossing coins, rolling dice, or selecting playing cards from a deck.

That pedagogic examples such as these, having well defined, quantifiable outcomes-spaces, do not provide an adequate basis for understanding *subjective* probabilities, where there is no well-defined outcomes-space, is self-evident.[2]

---

[1] I worded this sentence carefully to sidestep issues of the quantum nature of reality.

[2] There is no universal agreement among probability experts, of which I make no claim to club membership, as to what constitutes subjective probability, epistemic probability, etc. Readers should feel free to substitute their own favorite descriptor.

[ASIDE: The distinction between probability calculations where there is a well-defined set of outcomes, and the fascinating blend of judgment and mathematics that constitutes subjective probability, seems to be unknown to many journalists and radio producers. I and others with a public math identity frequently receive phone calls from media professionals who want to know the probability of some surprising, newsworthy occurrence. "Just what is the probability that a man would bite a dog?" would be the iconic question of this type, though I confess I have never actually been asked that one. I generally play along with the request, by providing a "Fermi-type" answer, but I long ago gave up any expectation that the published or broadcast version of my answer would be accompanied by a statement of the assumptions I made in arriving at the number I gave, or even that I made any assumptions at all. All the journalist wants is a number with a lot of zeros.]

Unfortunately, those quasi-experimental, pedagogic scenarios (tossing coins, rolling dice, selecting playing cards from a deck) lead to a belief that probabilities are empirical properties *of the scenario*, rather than a measure of *our knowledge of the outcomes*. It happens this way.

We (i.e., teachers of one kind or another) typically introduce the notion of probability by first getting students to toss coins, roll dice, or select cards from a deck, or whatever, and tabulate the results. The numbers thereby obtained are, of course, empirically determined facts about that specific sequence of actions. They are measures of the *frequencies* with which various kinds of outcomes *actually occurred.*

We go on to say that, absent the inconceivable event that the world will start to behave differently tomorrow from the way it has done throughout history, those frequencies provide a reliable numerical assessment of what is likely to happen when we perform the same action some time in the future. We say, for instance, that if we roll an honest die ten minutes from now, the probability it will land 6 up is 1/6 and the probability that it will land with an even number uppermost is 1/2.

That is certainly true. But what seems to get lost when we make this step is that the probability refers to our knowledge—to our expectation of what is likely to happen. There is, after all, no event to quantify until it occurs. There can be no empirical fact-of-the-matter about the outcome of a future event. The fact that we do not have definitive knowledge of future events is precisely what makes probability so useful in planning for the future. It is not the fact that the die roll is in the future that makes calculation of the probability useful; it is the fact that we do not have the information about the outcome. A future event for which the outcome is known in advance does not require a probability calculation.

The distinction becomes clearer if we perform the pedagogic experiment differently. Suppose we split the students into two groups. One group rolls the die and sees the outcome, the other group does not. I, as instructor, ask the second group to tell me how confident they are that the die landed 6-up. If they have been paying attention to the experimental work earlier, they will answer "One-sixth." (There are various ways I could frame this question, such as the odds they would want in order to place a bet on being right, etc.) But if I ask the first group the same question, and force them to answer, they will say "Zero" or "One." They can see how the die landed. For them, there is no uncertainty. (In fact, rather than answer "Zero" or

"One" they are more likely to say there is no need to calculate a probability, since they know the outcome.)

The distinction between the two groups is what they know about the outcome. Probabilities quantify not events but *our information* about events.

If someone in the first group were to whisper the outcome to a friend in the second, then that person would also answer "Zero" or "One" (or say they knew the actual outcome). Acquiring information can change the probability we assign to an event. In real life probability calculations, that is the norm!

If, when we introduce students to probabilities, we were more careful to ensure that they appreciated what exactly that number between 0 and 1 applies to, I think we would end up with less confusion than at present. In particular, a student who appreciates that probability quantifies *information* (or *statements*, if you prefer) will not find it surprising that the acquisition of more information can result in a revision of the probability we assign to some event.

For instance, many of the correspondents I have had exchanges with following articles I have written on Bayes' theorem and its various applications, have been unable to rid themselves of the belief that there has to be a "true" probability, some specific number that we just don't yet know.

Experts who have not had exchanges with interested, intelligent members of the public might be surprised to learn that a common belief about, say, DNA identification, is that its use is associated with a definite probability, somewhere strictly between 0 and 1, *that suspect X did commit the crime*. There is, of course, a (definitive) fact either that X did commit the crime or that X did not commit the crime. All my readers know that. But they also hold the belief that there is a definite "probability that X committed the crime"—a number strictly between 0 and 1—and that Bayes' theorem somehow provides a means of honing in on that number. In short, they think the probability is attached to the state of the world, rather than to our knowledge of the world.

The fact that they simultaneously hold that belief in an objective number strictly between 0 and 1 and the common-sense knowledge that a given individual either is or is not guilty (probability 0 or 1 if you have to mention one) only serves to emphasize the fundamental confusions probabilities generate.

To my mind, the frequentist approach to probability as an introductory pedagogy should be abandoned. True, it has the undoubted appeal of starting with student activity—rolling dice, tossing coins, etc. The trouble is that those are totally atypical examples. Practically every real-life application of probability theory occurs when there is no possibility of repeating a process many times and counting outcomes.

That is certainly the case with the two children problem. Terrorist attacks on airplanes are decidedly one-off events, yet our air travel is very visibly influenced by probability calculations. And weather is never repeated. Explanations couched in the form "Four times out of five when the conditions are similar to this it will rain" convince us as thought-experiments, not because they report actual outcomes. (Though such data actually is collected, and does form the basis of the computer programs the weather forecasters use.)

Even for the pedagogic examples of rolling dice and tossing coins, any actual experimentation is never more than motivational. We believe that an honest die will land 6-up 1/6th of the time not because we have rolled a die sufficiently often, but because we recognize the symmetry in the outcomes. The convincing evidence comes not from the physical experience, but from the thought-experiment that follows.

Given that almost every actual application of probability is in the context of a one-off event, introducing the concept through atypical classroom experiments probably does more harm than good. If we do continue to use them—and let me stress that I do see the value in starting with some experimental activities—then we need to make it clear that those are merely motivational, and that the real power of probability theory comes from reflection on the nature of events in the world and what we can know about them based on the information at our disposal.

Avoiding misconceptions about the objectivity of probabilities won't make probability an easier concept to master. It is inescapably problematic. That much becomes clear when you see the difficulty the famous French mathematician Blaise Pascal had coming to terms with Pierre de Fermat's solution to the Problem of the Points, which I described in some detail in my book *The Unfinished Game*.[3] But by exercising a bit more care in the beginning, we could, I think, eliminate a widespread confusion as to exactly what probabilities apply to, and thereby avoid the resistance many people have to the idea that probabilities can change as a result of acquiring more information! For in the real world, they can, and they usually do. For what a probability does is assigning a number to the aggregate of the currently available information.

Stanford University, USA Keith Devlin

[3]Keith Devlin, *The Unfinished Game: Pascal, Fermat, and the Seventeenth Century Letter that Made the World Modern*, Basic Books (2008).

Introduction to *Probabilistic Thinking: Presenting Plural Perspectives*

"Research into the development of probabilistic thinking... has occurred largely during the last 50 [now 60] years" (Jones and Thornton 2005, p. 65). Further, "the beginning of studies by researchers in mathematics education" (ibid.) did not occur until the 1970s and 1980s. Thus, research into probabilistic thinking in the field of mathematics education is a relatively recent endeavor.

The last 60 years of research have resulted in important pieces of literature. These pieces include, but in no way are restricted to: books (Piaget and Inhelder 1951/1975; Fischbein 1975); edited books (e.g., Kapadia and Borovcnik 1991; Jones 2005); dedicated yearbooks (e.g., Burrill and Elliot 2006; Shulte and Smart 1981); special journal issues (e.g., Borovcnik and Kapadia 2009; Biehler and Pratt, 2012); and research syntheses (e.g., Borovcnik and Peard 1996; Garfield and Ahlgren 1988; Hawkins and Kapadia 1984; Jones, Langrall and Mooney 2007; Shaughnessy 1992, 2003). As it is not our intention to summarize or synthesize existing research, which has elegantly been done elsewhere (including the very research we have highlighted), we recommend that you (re)visit the above literature. Instead, in this Introduction, we have chosen a more arduous task: to identify the next period of research for probabilistic thinking in mathematics education.

Emerging from the last 60 years, the (albeit brief) history of investigations into probabilistic thinking is ripe for analysis. Jones and Thornton (2005), in their "historical overview of research on [probabilistic thinking and] the learning and teaching of probability" (p. 66), classified the research into three phases: the Piagetian Period, the Post-Piagetian Period, and Contemporary Research. According to Jones and Thornton, Phase One, the Piagetian Period, is defined by the research of Jean Piaget (and Barb Inhelder), which occurred largely during the 1950s and 1960s. Jones and Thornton's second phase, occurring during the 1970s and 1980s, was denoted the Post-Piagetian Period for the strong Piagetian influence on the research being conducted. Also key to the Post-Piagetian Period was seminal research on intuition (e.g., Fischbein 1975), heuristics (e.g., Tversky and Kahneman 1974) and the beginning of studies conducted by researchers in the field of mathematics education. In keeping with two-decade increments, Phase Three, (the) Contemporary Research (Period), occurred during the 1990s and 2000s. This period, shaped by "the emergence of probability and statistics as a mainstream [curricular] strand" (p. 79), is defined by research investigating: curriculum, the teaching and learning of probability and learning environments.

Jones and Thornton's (2005) historical review was not without caveats. For example, they made clear that their review "concentrated more… on research that occurred prior to 1990" (p. 83) and that "they only provided an overview of contemporary research" (ibid.). They also noted: "although it is premature to evaluate historically the significance of probability research in this third and contemporary phase, it is clear that the volume and diversity of the research is greater than in the previous two phases" (ibid.). Nearly ten years on, we contend it is no longer premature to evaluate the historical significance of probability research conducted during the Contemporary Research Period. Further, in continuing with the established two-decade intervals, the 2010s, we also contend, should mark the ushering in of a new phase of research in probabilistic thinking.

As mentioned, Jones and Thornton's (2005) historical review focused more on the Piagetian Period and the Post-Piagetian Period, and provided an overview of Contemporary Research. However, a mere two years later, Jones et al. (2007) "synthesized research in the field, especially research that had been conducted since Shaughnessy's (1992) review" (p. 909). Alternatively stated, Jones et al. synthesized the volume and diversity of research that occurred during the Contemporary Research Period (as had been overviewed in 2005). From this analysis (despite Jones and Thornton's [2005] phase/period terminology not being adopted), one can evaluate the historical significance of the Contemporary Research Period. To this end, research into probabilistic thinking during the 1990s and 2000s, the Contemporary Research Period, is defined by research investigating the teaching and learning of probability in classrooms and schools, which is due, in large part, to probability becoming a mainstream strand of worldwide curricula.

The historical significance of the Contemporary Research Period can be judged according to research investigating probabilistic thinking and the teaching and learning of probability in classrooms and schools. However, as we will now demonstrate, the significance of the period may also be evaluated historically with respect to another phenomenon. This phenomenon, we further contend, will have a direct impact on the next phase of research into probabilistic thinking in mathematics education.

The Contemporary Research period can also be historically judged for the following phenomenon: liberal utilization of research topics in the field of mathematics education. In many instances (exceptions noted), mathematics education research topics, when utilized in research investigating probabilistic thinking, are used with (academic) impunity, that is, they are exempt from the academic scrutiny applied in other areas of research in the field of mathematics education.

Cementing this phenomenon into the annals of history is the (current) list of research topics in mathematics education that are liberally utilized in research investigating probabilistic thinking: beliefs, cognition, conceptions (and misconceptions), heuristics, intuition, knowledge, learning, modeling, reasoning, teaching, theory, thinking, and understanding. In addition, and perhaps even more surprisingly, there is also liberal usage of terms inherent to probability (e.g., risk, stochastics and subjective probability). Numerous examples, found throughout the Contemporary Research Period, support the claim that research topics in mathematics education are utilized without (a modicum of) reference to the extensive established bodies of literature that exist for each of the abovementioned topics in the field of mathematics education.

The currently accepted liberal utilization of research topics in mathematics education coupled with a relatively brief history of research (beginning in the 1970s) provides evidence that research investigating probabilistic thinking exists, currently, on the fringe of mathematics education. However, there are several indicators that probabilistic thinking, as a research topic, is slowly becoming a part of mainstream mathematics education. The recent increase in dedicated working groups and discussion groups at major mathematics education conferences, special journal issues, dedicated chapters in handbooks and yearbooks and, for that matter, the publication of *Probabilistic Thinking*: *Presenting Plural Perspectives* (as part of the *Advances in Mathematics Education Series*), are all indicators that the move to mainstream mathematics education is underway.

The age of the field, the predefined length of research phases and the indicators of a shift from the fringe to mainstream mathematics education signify the dawn of a new phase of research for probabilistic thinking in mathematics education. Naturally, the question of what to name the next period of research is raised. Of course, Post-Contemporary comes to mind. However, taking into account the move from the fringe to mainstream mathematics education and, as a result, the necessity to discontinue the liberal utilization of research in mathematics education during investigations into probabilistic thinking, we (prematurely) propose Phase Four be denoted the Assimilation Period. However, as you will encounter, *Probabilistic Thinking*: *Presenting Plural Perspectives* houses many different interesting developments. Any one (or a combination of) these developments may, at some point, better define the Assimilation Period. Perhaps the Assimilation Period will become known for (and named accordingly): the adoption of a unified approach to the teaching and learning of the classical, frequentist and subjective interpretations of probability; defining subjective probability; a (re)new(ed) interest in heuristics; a return to the field's psychological roots; or risk. Regardless of what the next phase is ultimately known for, we think there is a chance that *Probabilistic Thinking*: *Presenting Plural Perspectives* will play a role in the transition from the Contemporary Research Period to the Assimilation Period.

University of Saskatchewan, Canada — Egan J. Chernoff
The University of Montana, USA — Bharath Sriraman

References

Biehler, R., & Pratt, D. (Eds.) (2012). Probability in reasoning about data and risk [special issue]. *ZDM—The International Journal on Mathematics Education, 44*(7), 819–952.

Borovcnik, M., & Kapadia, R. (Eds.) (2009). Research and developments in probability education [special issue]. *International Electronic Journal of Mathematics Education, 4*(3). Retrieved from: http://www.iejme.com/032009/main.htm.

Borovcnik, M., & Peard, R. (1996). Probability. In A. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education* (Part 1, pp. 239–288). Dordrecht: Kluwer.

Burrill, G. F., & Elliott, P. C. (Eds.) (2006). *Thinking and reasoning with data and chance: sixty-eighth yearbook*. Reston: National Council of Teachers of Mathematics.

Fischbein, E. (1975). *The intuitive sources of probabilistic of probabilistic thinking in children*. Dordrecht: Reidel.

Garfield, J. B., & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: implication for research. *Journal for Research in Mathematics Education, 19*, 44–63.

Hawkins, A. S., & Kapadia, R. (1984). Children's conceptions of probability—a psychological and pedagogical review. *Educational Studies in Mathematics, 15*, 349–377.

Jones, G. A. (Ed.) (2005). *Exploring probability in school: challenges for teaching and learning*. New York: Springer.

Jones, G. A., Langrall, C. W., & Mooney, E. S. (2007). Research in probability: responding to classroom realties. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 909–955). New York: Macmillan.

Jones, G. A., & Thornton, C. A. (2005). An overview of research into the learning and teaching of probability. In G. A. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 65–92). New York: Springer.

Kapadia, R., & Borovcnik, M. (Eds.) (1991). *Chance encounters: probability in education*. Dordrecht: Kluwer.

Piaget, J., & Inhelder, B. (1975). *The origin of the idea of chance in students* (trans: L. Leake Jr., P. Burrell, & H. D. Fischbein). New York: Norton. (Original work published 1951.)

Shaughnessy, J. M. (1992). Research in probability and statistics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 465–494). New York: Macmillan.

Shaughnessy, J. M. (2003). Research on students' understanding of probability. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 216–226). Reston: National Council of Teachers of Mathematics.

Shulte, A. P., & Smart, J. R. (Eds.) (1981). *Teaching statistics and probability: 1981 yearbook*. Reston: National Council of Teachers of Mathematics.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124–1131.

Contents

Perspective I: Mathematics and Philosophy

Preface to Perspective I: Mathematics and Philosophy

Egan J. Chernoff and Gale L. Russell

> The theory of probability has a mathematical aspect and a foundational or philosophical aspect. There is a remarkable contrast between the two. While an almost complete consensus and agreement exists about the mathematics, there is a wide divergence of opinions about the philosophy. (Gillies 2000, p. 1)

Within the wide divergence of opinions about the philosophy of probability, there is one significant bifurcation that has been recurrently acknowledged since the emergence of probability around 1600 (Hacking 1975). Hacking describes this duality of probability as the "Janus-faced nature" (p. 12) of probability, explaining "on the one side it is statistical, concerning itself with stochastic laws of chance processes" (ibid.); and "on the other side it is epistemological, dedicated to assessing reasonable degrees of belief in propositions quite devoid of statistical background" (ibid.). Although the phrase "Janus-faced" continues to be used throughout probability (related) literature, the terms used to describe the two different faces (i.e., the different theories or interpretations) of probability have not been similarly adopted.

In fact, the nomenclature associated with the duality of probability is (for various reasons) inconsistent. For example, Hacking (1975) used the terms "aleatory" and "epistemic" for the two faces. Gillies (2000), while adopting Hacking's epistemic, found the term aleatory "unsatisfactory" (p. 20) preferring the term "objective" instead. Although different terms (all with their own specific issues) have been used by different individuals to describe the duality of probability, the terms that (perhaps) most familiarly represent the Janus-faced nature of probability are "Bayesian" and "frequentist" or, respectively, "subjective" and "objective." (See McGrayne 2011 for a detailed account of the history of the Bayesians and the frequentists.)

Arguably, the theory of probability, currently, has four (not two) primary interpretations. Whereas "logical" and "subjective" denote edges of the division within subjective or Bayesian (or "belief-type" Hacking 2001) theory, "frequency" and "propensity" denote the extremes of the division within the objective or frequentist

E.J. Chernoff (✉) · G.L. Russell
University of Saskatchewan, Saskatoon, Saskatchewan, Canada
e-mail: egan.chernoff@usask.ca

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_1,

(or "frequency-type" Hacking 2001) theory. Worthy of note, issues surrounding the terms used for these four main interpretations persist. For example, Gillies (2000) notes that "subjective" is used, concurrently, as a general classifier and as a specific theory. Resolutions to this issue include Hacking's (2001) use of the term "personal" instead of "subjective" when discussing specific theory. Regardless of the terminology used, however, the four primary interpretations in probability theory are not likewise found in the field of mathematics education.

In the field of mathematics education, only three (currently) dominant philosophical interpretations of probability are present: classical, frequentist and subjective. Akin to probability theory, nomenclatural issues (e.g., inconsistent use of multiple terms) also exist. The classical interpretation of probability in mathematics education is known alternatively as "a priori" or "theoretical;" whereas the frequentist interpretation of probability is differently known as: "a posteriori," "experimental," "empirical" or "objective." Through informal consensus, despite the multitude of terms, when an individual uses the term "classical probability" or "frequentist probability," there is little confusion. Unfortunately, the same cannot be said for the "subjective probability."

Chernoff (2008), in an examination of the state of probability theory specific to the field of mathematics education, concluded that the state of the term "subjective probability" is *subjective*. There is, once again, inconsistent use of multiple terms (e.g., "subjective," "Bayesian," "intuitive," "personal," "individual," "epistemic," "belief-type," "epistemological" and others). Moreover, Chernoff also determined that the term subjective probability in mathematics education has dual meaning—"subjective probability" is concurrently used as a general classifier and as a specific theory. In an attempt to rectify the dual usage of the term, Chernoff suggested subjective probability remain the general classifier and a further distinction be made, more aligned with the distinction found in probability theory, for the specific theory. In further examining parallels between mathematics education and probability theory, Chernoff contended that subjective probability (the specific theory not the general classifier) in the field of mathematics education trended toward the "personal" (or "subjective") interpretation found in probability theory rather than the "logical" interpretation. As a result he suggested that "subjective probability" be used as a general classifier while "personal probability" (or other terms) could be used for the specific theory.

In a recent, comprehensive synthesis of probability research in mathematics education, Jones et al. (2007), declared that "it is timely for researchers in mathematics education to examine subjective probability and the way that students conceptualize it" (p. 947). Their declaration—(perhaps) made in part because the authors (i) "were not able to locate cognitive research on the subjective approach to probability" (p. 925) and (ii) found that subjective probability "is not widely represented in mathematics curricula" (p. 947)—is well received. However, it must be noted, the nomenclatural issues inherent to subjective probability, documented above, remain unresolved as the subjective interpretation makes its way into curricula, classrooms and mathematics education research around the globe.

Mathematics education and probability theory are similar, in that the issues associated with differing philosophical interpretations are found in both domains; however, the domains differ with respect to the issues surrounding the infamous feud between the Bayesians and the frequentists. A close read of important pieces of literature in mathematics education reveals individuals' support for particular interpretations of probability (e.g., Hawkins and Kapadia 1984; Shaughnessy 1992). Despite these declarations of affinity for one interpretation over another, the feud between Bayesians and frequentists does not (appear) to exist in the field of mathematics education. The most likely reason that this feud is not found in mathematics education is because the very same research that advocates one interpretation over another also champions an approach to the teaching and learning of probability that "utilize[s] subjective approaches in addition to the traditional 'a priori' and frequentist notions" (Hawkins and Kapadia 1984) or, alternatively stated, "involves modeling several connections of probability" (Shaughnessy 1992, p. 469). In more general terms, research in mathematics education continues to advocate for "a more unified development of the classical, frequentist, and subjective approaches to probability" (Jones et al. 2007, p. 949).

As you will encounter, the chapters in this perspective all contribute to a more unified development of the three different approaches to probability in mathematics education, albeit in different ways. Through historical accounts of mathematics and philosophy, discussion of the importance of puzzles and paradoxes as a bridge between philosophy and mathematics and research focused on modeling and technology, these chapters may ultimately help align the interpretations of probability in mathematics education with those found in probability theory.

References

Chernoff, E. J. (2008). The state of probability measurement in mathematics education: a first approximation. *Philosophy of Mathematics Education Journal*, *23*. Retrieved from http://people.exeter.ac.uk/PErnest/pome23/index.htm.

Gillies, D. (2000). *Philosophical theories of probability*. New York: Routledge.

Hacking, I. (1975). *The emergence of probability: a philosophical study of early ideas about probability induction and statistical inference*. Cambridge: Cambridge University Press.

Hacking, I. (2001). *An introduction to probability and inductive logic*. Cambridge: Cambridge University Press.

Hawkins, A. S., & Kapadia, R. (1984). Children's conceptions of probability—a psychological and pedagogical review. *Educational Studies in Mathematics*, *15*, 349–377.

Jones, G. A., Langrall, C. W., & Mooney, E. S. (2007). Research in probability: responding to classroom realties. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 909–955). New York: Macmillan.

McGrayne, S. B. (2011). *The theory that would not die: how Bayes' rule cracked the Enigma code, hundeted down Russian submarines and emerged triumphant from two centuries of controversy*. New Haven: Yale University Press.

Shaughnessy, J. M. (1992). Research in probability and statistics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 465–494). New York: Macmillan.

A Historical and Philosophical Perspective on Probability

Manfred Borovcnik and Ramesh Kapadia

Abstract This chapter presents a twenty first century historical and philosophical perspective on probability, related to the teaching of probability. It is important to remember the historical development as it provides pointers to be taken into account in developing a modern curriculum in teaching probability at all levels. We include some elements relating to continuous as well as discrete distributions. Starting with initial ideas of chance two millennia ago, we move on to the correspondence of Pascal and Fermat, and insurance against risk. Two centuries of debate and discussion led to the key fundamental ideas; the twentieth century saw the climax of the axiomatic approach from Kolmogorov.

Philosophical difficulties have been prevalent in probability since its inception, especially since the idea requires modelling—probability is not an inherent property of an event, but is based on the underlying model chosen. Hence the arguments about the philosophical basis of probability have still not been fully resolved. The three main theories (APT, FQT, and SJT) are described, relating to the symmetric, frequentist, and subjectivist approaches. These philosophical ideas are key to developing teaching content and methodology. Probabilistic concepts are closer to a consistent way of thinking about the world rather than describing the world in a consistent manner, which seems paradoxical, and can only be resolved by a careful analysis.

1 Introduction and Sources

The short history of probability as compared to other mathematical concepts has deep reasons, which differ from those that caused the axiomatic treatment of numbers to occur so much later than Euclid's axiomatization of geometry. One reason behind this tardiness may be the confusion about the purpose of the concepts. Will probability in a mathematical formulation help us to know the future in advance? Or, is probability just a view on the world by which one might get an advantage in

M. Borovcnik (✉) · R. Kapadia
Institut für Statistik, Alpen-Adria Universität Klagenfurt, Universitätsstraße 65, 9020 Klagenfurt, Austria
e-mail: manfred.borovcnik@uni-klu.ac.at

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_2,

acting more consistently? The positive answer to the second question is of limited value to the emotive hope of the first question.

For example, the statement that the probability to roll a six with a die is $1/6$ arises from the application of a mathematical model. To ask whether a die has a physical property to show a six (be it one sixth or an unknown number p) reveals a philosophical bluntness, as if such a property is comparable intrinsically to the weight or the diameter of a coin. In a general philosophical debate, a probability statement as a physical property of the die would be construed to be similar, yet different from the property of the 1 US cent coin to weigh 2.05 gram and have a diameter of 19.05 mm. One major difference is that a probability statement of the die describes a model property, which might be indirectly measured by some ideal procedure but not measured in reality. One way to decide whether to adopt a probabilistic view on a specific die is to investigate the results of tossing it.

Though a sound mathematical foundation was published in 1933, this has not clarified the nature of probability. There are still a number of quite distinctive philosophical approaches which arouse controversy to this day.

There are useful sources for the history of probability like Todhunter (1865), David (1962, early traces), Maistrov (1974, not much before Pascal and Fermat, with a special focus on the Russian school till von Mises), Schneider (1989, till Kolmogorov, in German), Porter (1986, frequentism in the nineteenth century), Daston (1988, classical probability in the enlightenment of the eighteenth century), Hald (1990, 2007), Stigler (1986, till 1900). Philosophical issues are dealt with in Stegmüller (1973), Fine (1973), Barnett (1973), Weatherford (1982), and Hacking (1975, 1990). There are many books on probability itself such as von Plato (1994) and Çinlar (2011). The interested reader should consult these books for more detailed coverage of the ideas underlying probability. We have used these books to source key ideas which are relevant in probability education. Readers should also note that we have used these references rather than the originals, many of which are obscure and in other languages.

For our purpose of teaching probability, we will focus on the following three philosophical theories, which are discussed in more detail later:

- Classical a priori theory (APT); this is developed from the original classical theory.
- Frequentist theory (FQT); this is based on experimental results.
- Subjectivist theory (SJT); this is based on personal belief and linked to Bayes' theorem.

2 From Divination to Combinatorial Multiplicity

2.1 Early Origins in Divination and Religion

Early notions of chance can be found in the ancient cultures of Indians, Babylonians, and Egyptians. David (1962) refers to the astragalus—a bone in the heel of a sheep—as the earliest known object used for games of chance around 3500 B.C. It is possible that primitive dice were constructed by rubbing the round sides of the astragalus until it was approximately flat. Yet, so long as the real bones were used, their natural variation would affect their comparability and thus increase their characteristic to be unpredictable, a key element of chance. The cubes in Babylon of 3000 B.C. made from well-fired pottery were nearly symmetrical dice.

However, there were no systematic investigations—neither of combinatorial multiplicities nor of frequencies to study the results. In her book on the origins and history of probabilistic ideas from ancient times to the Newtonian era, David (1962, p. 21) speculates on why the conceptual progress of probability was so tardy:

> It was fifteen hundred years after the idealization of the solid figure before we have the first stirrings, and four hundred and fifty years after that before there was the final breakaway from the games of chance, if indeed we have really accomplished it even today.

In fact, there were some attempts to judge the likelihood of the outcomes as Cicero in his *De divinatione* (44 B.C.) expresses in the following argument:

> They are entirely fortuitous you say? Come! Come! Do you really mean that? ... When the four dice produce the venus-throw you may talk of accident: but suppose you made a hundred casts and the venus-throw appeared a hundred times; could you call that accidental? (David 1962, p. 24).

In one throw, it is unlikely to have all four sides up with a different result (the venus-throw), and it is even more unlikely to get this result a hundred times in series. The argument was based on the near impossibility of that outcome (not on a combinatorial basis) but in the sense of seeing god's will in such an outcome. Cicero continues to speculate on the character of randomness:

> Is it possible, then, for any man to apprehend in advance occurrences for which no cause or reason can be assigned? (David 1962, p. 24)

There are sound reasons that theoretical arguments were not developed. The Greek philosophy had an ideal of true relations, which prevented the construction of theoretical hypotheses from empirical data. The Greek tradition is illustrated by the "beauty" of Platonic bodies, which have underlying symmetries, and Euclid's geometry, which is based on axioms. Data about variability was ignored as contrary to their ideal of scientific argument. This is an external explanation of the tardy development of concepts for probability. An internal one might be that probability right from the early origins was intimately bound to divination to predict the future (Borovcnik 2011).

Divine judgement at religious ceremonies (at Delphi and elsewhere) was often based on what we would now call games of chance—beans were drawn from a bowl of white and black beans to answer binary questions. Four or five astragali were thrown and the answer to a question was linked to the combinatorial possibilities: for each outcome a structural response to the posed question was prepared. The combinatorial multiplicity may have been implicitly known but links to the outcome were not investigated.

Dice were recognized as a tool to decide fairly or to explore god's will for a decision. From the letters of Lucas we see that—when they could not decide between two persons—they cast lots for them. From the Old Testament there are several places known where a decision was determined by a chance experiment (about 70 places are known to refer to such a use of games of chance, see Logos Bible Software n.d.). For exploring god's will it was not so important that lots were fair.

In the Christian era, starting with the acceptance of Christianity as the only allowed religion (under Theodosius, 380 A.D.), games of chance lost prestige as everything that happens is determined by the will of god. This attitude is expressed by St Augustine (354–430) as we may see from a quotation from David (1962, p. 26):

> [...] nothing happened by chance, everything being minutely controlled by the will of God. If events appear to occur at random, that is because of the ignorance of man and not in the nature of the events.

Such a strict attitude of determination by god is hardly reconcilable with the free will of an individual. Thus Christian philosophy hindered theoretical discussion of randomness.

It is interesting that games of chance did not remain the exclusive preserve of priests; they became an important leisure activity throughout the Roman Empire. The addiction to play such games of chance may nourish a deep desire to explore god's will in everyday situations (the popularity of games of chance is unbroken till today—the business of games of chance is an important branch of applications of probability). At times, Roman emperors forbade games with no success.

Considerable experience would have been gained by priests and other important people from casting dice or drawing beans out of urns for divine judgement at religious ceremonies. In some ways, it is curious that a conceptual breakthrough failed to appear to formalise probability, based on the regularity of the fall of dice. It could be that priests were taught to manipulate the fall of the dice to achieve a desired result, as the interpretation of divine intent, conveyed through their pronouncements. More critically, speculation on such a subject might have brought a charge of impiety in the attempt to penetrate the mysteries of the deity. Moreover, men knew that nature was fickle and while there were elements of regularity such as the rising of the sun every day many other elements of life were hard to predict, not least the weather. Divination was important and a key element of the power of religion and its high priests, as a means of control of ordinary people.

According to Borovcnik et al. (1991, p. 28), one should not overestimate this tardiness of conceptualization as this is also true for other disciplines.

> The development of a causal physical approach to science instead of a deistic one was marked by great controversies, even if today one finds such ideas naturally accepted. Think of Galilei's (1564–1642) troubles with the Pope. Euclidean geometry was not the breakthrough as is generally viewed, neither in axiomatic thinking nor in geometry. It only provided rules for construction but no formal concepts; the status of the parallel axiom was not clarified until the work of Gauss and Lobachevski in the nineteenth century; the concept of continuity of lines and planes was developed even later, though it is implicitly assumed by Euclid. In arithmetic, on the other hand, no sound answer was given to the question of axiomatizing numbers until Peano over 100 years ago. However, a comparable milestone was not reached in probability until 1933.

One further obstacle to conceptual progress might also lie in the following aspect, which focuses on the individual faced with a random situation, be it ruled either by coins or by divine judgement: nothing but the *next* explicit outcome counts. However, this cannot be predicted by any human being. Conceptual progress can be achieved by interpreting this single case as one representative of a series of future (or hypothetical outcomes). Analytical philosophy of science has not come up with any satisfying solution of that problem (as expressed by the comprehensive and extensive research of Stegmüller 1973). But, all the more, for the individual any solution to the series of future outcomes has no attraction as—even if there were more comparable situations to face in future—the individual would always focus on the present situation with all its ingredients including fate, god's decision, impact of outcomes, which will never be the same. In their experiments, Kahneman and Tversky (1979) have shown that even an interchange in gain or loss in a situation is very influential. Thus, the answer to the series is never an answer to the single case. Surprisingly enough, a probability for the series still yields some information for the single case. So, the knowledge about tossing a coin allows a fair decision between two opponents when deciding the sides to play in a football game.

2.2 Emergence of the Rule of Favourable to Possible: Combinatorial Multiplicity

The poem *De vetula* from the thirteenth century may be the first explicit link between combinatorial multiplicity and frequencies of outcomes. For three dice, all 216 possible cases are listed. Of the 16 possible sums, it is advisable to put wagers according to expected gain as "you will learn full well how great a gain or loss any of them is able to be" (Bellhouse 2000, p. 135, cited from Batanero et al. 2005, p. 20); however, this and Peverone's attempt to enumerate the possibilities of a game (Peverone 1558) do not establish the notion of probability. Similarly, Cardano's deliberations in his *Liber de ludo aleae* (approx. 1564) confuse expected value and probability. Cardano discusses how equally easy it is to get any of the numbers in casting one die but then proceeds with an awkward argument about the consequences:

> One-half the total number of faces always represents equality; thus the chances are equal that a given point will turn up in three throws, for the total circuit is completed in six, or again that one of three given points will turn up in one throw. [...] The wagers therefore are laid in accordance with this equality if the die is honest. (Cardano approx. 1564, quoted from David 1962, p. 58)

It is doubtful that Cardano saw a clear relation between empirical frequencies and a theoretical concept like probability.

The main books on the history of mathematics give the credit of the first conceptual approach to probability to a famous correspondence. This suits a romanticised viewpoint, even if the authenticity can be doubted. It is the evil of gambling which seemingly led to the fall of determinism. More generally, people do build up a personal view of chances from daily life and its encounters with chance. Thus it was a century after Cardano that Pascal and Fermat made progress in conceptualizing probability in their exchange of letters in 1654 (published in Fermat 1679; see also David 1962). They discussed and solved two specific problems, de Méré's problem and the Division of Stakes (problem of points). Both problems, which are described below, relate to gambling and fairness, using the idea of proportion, which is the basic building block of probability; yet the use of proportion is not always as straightforward as one might imagine. This is too often neglected when linking these ideas within the educational sphere.

De Méré's Problem Two games are compared. In game 1, the player wins if there is at least one 'six' in four throws of a die; in game 2, the player wins if there is at least one 'double six' in 24 throws of two dice. At first sight, the games seem to be very similar. The solution of Pascal and Fermat is based on an exhaustive enumeration of the sample space (or, fundamental probability set), following the work of Galilei (who took up the three dice of *De vetula*):

$$\text{P(win in game 1)} = 1 - (5/6)^4 = 671/1296 = 0.518 > 1/2,$$
$$\text{P(win in game 2)} = 1 - (35/36)^{24} = 0.491 < 1/2.$$

According to Ore (1953, p. 411), de Méré rejected that solution and referred to the equity of proportions: 24 (opportunities to get the desired result, which were seen as favourable cases) to 36 (possible cases) has the same proportion as 4 (opportunities) to 6 (possibilities).

It is romantic folklore that de Méré won a fortune by game 1 and lost everything by game 2. However, a deviation of 0.009 from 0.5 in game 2 requires roughly 13,000 data to detect it (depending on the significance level of a modern statistical test). Second, with the experience of 80 bets per week (roughly 4 nights of gambling with 2 hours on the dice table) a 95% confidence interval for the final balance in game 2 for half a year would be roughly -122 to 53 with an expected value of -34 units. Multiply that with stakes of 100 instead of 1 unit still would result in expected losses of 3,400 and a—probabilistic—upper boundary for the losses of 12,200. A person who can stake 100 per game would not be deprived of all his fortunes if losing 12,000 within half a year.

De Méré's rhetorical rejection of the solution reflects a careless use of the "favourable to possible" rule developed at that time. Indeed, the favourable "cases" refer to repetitions of the experiment and are not a subset of the possible cases. There was a conflict between an enumeration of the sample space and the rule of favourable to possible, which still had to be clarified. Another source of complication might again be a confusion between probability and expected value. While the probabilities are different for both games, the expected number of the desired result is the same in both games—there are 4 trials, each has the expected value of $1/6$, thus the expected number of sixes in game 1 equals $2/3$. In game 2, with 24 trials with expected value of $1/36$ each, the expected number of times a double six occurs, is again $2/3 = 24/36$. We will illustrate the difference of both games by a *scenario* of 100 games (see the table below): in 48 runs of game 1 no six occurred and the bet is lost; for game 2, 51 bets are lost as no double six occurred.

In both series the average number of sixes (double sixes) is 66, which corresponds to the expected value of 2/3 per run of the game. While the bet is lost in game 2 more often, in game 2 the event "three or more" occurred more often than in game 1. However, there was no extra payment for this excess of double sixes.

How often	6's in game 1	66's in game 2
0	48	51
1	39	35
2	12	11
3 or more	1	3
Average	66	66
Win the bet	52	49

Division of Stakes If Peter and Paul are competing, how should they share the stakes if the series is ended at a time when Peter needs two points and Paul needs three more points to win? Suppose, for example, in a series of 11 Peter has 4 points while Paul has 3 points.

Huygens (1657) incorporated the problem and its probabilistic solution in his *De ratiociniis in ludo aleae*. The perception of the situation with probabilities, once introduced was so convincing that other arguments were defeated. The conclusion is that the stakes should be divided proportionately to the probability of winning in the continuation of games and neither in the proportions 3:2 nor 4:3 (this solution was favoured by Pacioli 1494). Pascal later developed his famous arithmetical triangle as a general method to solve similar problems involving binomial experiments.

Pascal and Fermat's approach sheds light on the correct application of what they termed to be the 'favourable to possible rule', but they made less progress in trying to formally define the concept of probability. They used probability pragmatically as the equal likelihood of outcomes in games of chance, which seemed to be intuitively obvious to them. Hence the emergence of the classical a priori theory (APT) of probability, which later was linked to the principle of indifference discussed below, based on the ideas of Laplace. Games of chance eventually served as a vital link between intuition and developing concepts as well as a tool to structure real phenomena. This view is also supported by Maistrov (1974, p. 48):

> These games did serve as convenient and readily understandable scheme for handy illustration of various probabilistic propositions.

3 Huygens, Bernoulli, and Bayes: The Art of Conjecturing

3.1 Expectation and Probability

A multi-faceted personality in the history of probability is Huygens (born in the Hague in 1629)—the real begetter, the man who synthesized ideas on *probability* in a systematic way. His book *De ratiociniis in aleae ludo* was published in 1657 and was not superseded for over half a century. But strangely his 14 propositions do not contain the word of probability, though he does mention chance. Huygens refers to situations of equal chances and later only to having p chances. And he derives no specific probabilities but proportions of stakes or a price of the situation.

> Proposition 3: to have p chances of obtaining a and q of obtaining b, chances being equal, is worth $\frac{pa+qb}{p+q}$.

In his analysis of the historical development, Shafer (1996, p. 16) explicitly refers to the non-probabilistic embedding of the emergent concepts:

> Pascal, Fermat, and Huygens were concerned with a problem of equity, not a problem of probability. They were pricing gambles, not evaluating evidence or argument. But it did not take people long to draw the analogy.

For Huygens there are two worlds with no direct connection. From games of chance with equally likely outcomes he derives the value of an enterprise (its expectation) as an economic price, a theory of equity. Huygens does not use the word probability to denote the proportion of stakes for a player as his probability of winning. In this approach, Huygens derives recursive rules for expected values if the basic situation is interpreted with probabilities from today's perspective. But he does not use the term expected value (a notion that came into the Latin translation by Francis van Schooten); instead Huygens speaks about the price or true value of a pay-off table.

Probability was also the ingredient of the other "theory" on decisions, on which Huygens worked. Probability was not yet a number but the collection of arguments (pro and con) in order to weigh these arguments. However, since Huygens, the frequency of specific observations becomes a possibility to substitute for other arguments:

> [...] especially when there is a great number of them [...] one can imagine and foresee new phenomena, which ought to follow from the hypotheses [...] and when one finds that therein the fact corresponds to our prevision. (Huygens in Treatise on Light, cited from Shafer 1996, p. 18)

Another application of probability emerged at that time. Huygens established mortality tables and treated frequencies in the same way as probabilities. Moreover, he defined theoretical concepts like the mean life time. These are the first signs of the idea of probability as a concept or theory based on relative frequency (FQT).

However, there was no deeper theoretical argument for such a statement before Bernoulli. Probability plays the role of an undefined elementary concept, which is derived from games of chance. Huygens does not develop any combinatorial methods which are consistent with his view on probabilities which are reduced to equity of cases of the underlying situation; his concept of expectation, recursively applied, yields solutions to various problems.

The first Englishman to calculate empirical probabilities was Graunt (1662). This was a natural extension of the Domesday Book, the most remarkable administrative document of its age, which enabled the King to know the landed wealth of his entire realm, how it was peopled and with what sort of men, what their rights were and how much they were worth. Speculations and research into probability theory did not concern the English, whose interest was in concrete facts.

Graunt's importance both as a statistician and an empirical probabilist lies in his attempts to enumerate as a fundamental probability set the population of London at risk to the several diseases such as are given in the Bills of Mortality recorded by London clerks. This gave further impetus to the collection of vital statistics and life tables. As Graunt epitomizes, the empirical approach of the English to probability was not through the gaming table, but through the more practical and raw material of experience. Graunt continues a tradition already started by Huygens to calculate empirical frequencies and deal with them as if they were probabilities without theoretical justification. For such problems like annuities, a practical solution was needed. A more theoretical problem was to estimate the mortality of diseases. These ideas are the precursors to insurance and risk.

Jakob Bernoulli (1713/1987) culminates this early development when he explicitly uses probability as parts of unity and supports this perception with his extensive work on combinatorics, in which the numerical probabilities of many standard problems (of the time) have been derived. He did not refer to proportions of stakes in games, bets, or economic enterprises, but used probability as a proportion or part of certainty for occurrence. His second achievement is what is now called the law of large numbers; Bernoulli himself called it *theorema aureum*, in which he derives mathematically a relation (a kind of convergence) between equal probabilities and the observed frequencies in the repetition of such games. It is this relation which justifies a more general concept of probability based on observed frequencies. Before this theorem, probability was derived from equal possibilities; after this theorem was proved, the way was open for a new concept of probability, namely probability linked to frequencies, which frees probability from equal likelihood.

Beyond this achievement, Bernoulli extended arguments on how to combine various parts of evidence on the basis of what we would now call independent events. His *Ars conjectandi* (1713)—the art of conjecturing—was published only years after his early death. Though Bernoulli thoroughly explored how to combine the evidence from several events, he did not develop realistic examples to illustrate the conditions that statements have to fulfil in order to fit his theorem. The use of relative frequencies can fail if the conditions of the single experiments are different. This was embedded and hidden in the Bernoulli distribution, which was the basis of his theorem, but was not noticed by the research community who had already gone further.

Bernoulli's philosophical ideas still stick to metaphysical determinism wherein everything is ruled by deterministic laws—be it weather, dice or eclipses of planets. Most phenomena were perceived to be so complex that it would be pointless to study the possible cases. However, his law of large numbers allows the use of observed frequencies for probability:

> [...] to ascertain it from the results observed in numerous similar instances. (Maistrov 1974, p. 69)

3.2 Obstacles and Further Developments

One often learns the most about a concept by exploring situations, in which it fails. Bernoulli had given one possibility to connect probability to relative frequencies but this was only part of his approach. He knew that this could fail either with insufficient data to measure the probability accurately (probability was used as a physical quantity at that time) or because the condition of independence is not satisfied.

Bernoulli's introduction of the new term probability, as being a part of certainty (i.e. a fraction between 0 and 1 instead of proportions of favourable to unfavourable cases) was generally accepted. However, the delay of the publication of his work allowed the community to develop ideas along other lines ignoring his approach that probability could be measured by empirical frequencies with enough data and the relation between frequencies and probabilities resulting from his *theorema aureum*. Thus, de Moivre's *Doctrine of chance* (1718) had a cursory reference to the law of large numbers and only in its second edition of 1738/1967, was the theorem included as a mathematical highlight, yet completely isolated from the rest of the book. Bernoulli's art of conjecturing to combine probabilistic arguments was abandoned, and when Bayes (1763) went back to solve it, his approach was very different. But before discussing Bayes, we describe a key aspect of Huygens' expected value of a situation, which remained an ingredient—the St Petersburg problem.

Cramer corresponded with the Bernoulli family about the following problem, which was published in 1738 as an appendix by Daniel Bernoulli (1738/1954) in the Petersburg Commentaries.

St Petersburg Paradox Two players A and B toss a coin until it shows a 'head' for the first time. If this occurs at the nth trial, then player B pays 2^{n-1} écus to player A. What amount should A pay to B before this game starts in order to make it fair?

If X denotes the amount B pays, then its sample space is (in theory) all natural numbers. The expected gain is $E(X)$, but this is infinite as the series involved diverges:

$$E(X) = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + \cdots + 2^{n-1} \cdot \frac{1}{2^n} + \cdots.$$

Thus player A would have to pay an infinite amount of money to B before the game starts.

A modern standpoint would see no problem. If an expected value is infinite this is seen as trivial since a mathematical model of a situation cannot reflect all its properties. The intuitive clash is irrelevant that we cannot play a game for ever. If, in a random experiment, the model used leads to a random variable with no existing expected value, everything is different. The Bernoulli law of "convergence" of the relative frequencies to the underlying probability is no longer valid if the expected value does not exist.

However, we return to the historic situation of the curiosities the problem induced. No one can play such a game as no person has an infinite amount of money.

Yet the expected length of the game is 2 and the associated payoff is 2. We may even be intuitively ready to stake 10 units (historic unit was écu). How can Huygens' expected value deliver a value of infinity in what seems to be a normal situation? At the time, people did not think they were applying a mathematical model to a real situation, or that the model could be inadequate.

Their approach sheds light on the historic perception of probability in the eighteenth century. Probability was not yet anchored by a unified theory, nor was Bernoulli's theorem common-place. Probability was perceived as kind of provability. So the mathematical consequence of getting an infinite value of a game was unacceptable. Much later, Venn formulated a harsh critique. He asserts that no man would pay 50 units as

> neither he nor those who are to pay would be likely to live long enough to obtain throws to remunerate him for one-tenth of his outlay, to say nothing of his trouble or loss of time. (Venn 1888, p. 155)

Venn did not refer to the waiting time until one wins an amount higher than 50 units (an event that has a probability of $1/2^6$). He meant, it would take more than a life to reach a point of time when the net balance of all previous games is positive (or at least zero).

One way to repair the theory is to introduce utility of money to make the expected value finite. A second way out of the dilemma was to introduce moral probability, a magnitude below which any probabilities are ignored. While the first leads to modern approaches of utility it does not solve the paradoxical feature of the Petersburg problem as one could change the payments of the game and still find an infinite expected utility. The time was not ripe to consider the special choice of the utility function merely as one of several possible models. The second approach shifts the dilemma to another level as to when a probability is small enough to be discarded and set to 0. The concept of moral expectation gained much support from Buffon, Condorcet, Laplace, and Poisson. We still have not solved such problems in modern inferential statistics where we introduce a significance level that resembles the moral probability of those times.

With regards to the utility of money, the richer one is, the less it makes sense for the person to gain more money in this way. Daniel Bernoulli (1738/1954) suggested that the utility of money should be a logarithmic function. With this in mind, the expected utility of the game would be

$$E\big(u(X)\big) = E\big(\ln(X)\big) = E\left(\sum_{n=1}^{\infty} \ln\big(2^{n-1}\big)\frac{1}{2^n}\right) = \ln 2,$$

but this would not reflect the approach properly as with utility, the additional gain of the game would add to the riches r already owned. Thus an equation like

$$\ln(r) = E\big(\ln(r - s + X)\big) = E\left(\sum_{n=1}^{\infty} \ln\big(r - s + 2^{n-1}\big)\frac{1}{2^n}\right)$$

has to be solved. In this equation, s is the stake required to play the Petersburg game. If the relation above turns to an equation, then fair stakes are achieved. A numerical solution yields $s \approx 9$ for $r \approx 50{,}000$, which is quite close to that amount of money a lot of people would be prepared to stake in the game. Such a concept was called *moral expectation* but it depends on the riches at the beginning and on the utility function used.

To use utilities instead of money has been discussed at various periods and is part of economic models today and an element of decision theory but utilities are not seen as part of the foundations of probability. Some of the phenomena detected by Kahneman and Tversky (1979) also have their origin in that people do use utilities implicitly. However, the utility function differs between people and is very subjective. In newer attempts to cope with risks, utility is revived via the term impact. The search for an objective utility function has been in vain though it led to some empirical laws of decreasing impact like the Webner–Fechner law (see Stigler 1986) in physiology.

The second way is to introduce the idea of a moral probability that should be neglected. Several authors advocate differing rules for the size of such an entity. Buffon (1777) argued—in comparison to the probability of a 56 year old dying within 24 hours of $1/10{,}189$ (which he derived from mortality tables)—that probabilities smaller than 10^{-4} should be ignored as no one actually fears that. This would lead to neglecting all terms in the expectation with $n > 13$ and yield a corrected expected value of 6.5.

In the course of his likelihood argument for the solar system, Buffon required a rule of thumb for deciding when a specific likelihood is small enough to establish the 'proof' of a common cause. For practical purposes, he proposed that an event with a probability below 0.0001 should be considered impossible and its complement as certain. Such ambiguous rules about *moral certainty* can also be found with other writers like Cournot. More recently, Borel (1943) also argued that we should deal with small probabilities as if they were zero: A human being should neglect 10^{-6}, in the history of earth, 10^{-15} and, for the cosmos, 10^{-50} should be neglected. For a human being, terms with $n > 19$ would therefore be discarded and an expected value of 9.5 in the St Petersburg puzzle would result.

To develop the first "significance test" similar thoughts were applied by Arbuthnot (1710–1712) in his "proof" of a divine order of gender: the probability is very small that for 80 successive years more males than females are born, so the hypothesis that the gender proportion is equal has to be rejected as it would produce an observation with a probability of $(1/2)^{80} \approx 10^{-25}$ (Gigerenzer 2004, traces the early origins of statistical inference). Therefore, the alternative hypothesis must hold (probability for boys greater than for girls), which Arbuthnot interpreted as an expression of a divine order. Once we decide the limit for the moral probability, any hypothesis that gives the observed event a probability smaller than this limit is probabilistically *disproved*.

Bayes' Formula and Inverse Probabilities Bayes (1763) solved the following problem, which we will describe in modern terms—in its discrete version, the Bayesian formula as it is called now: Let $H_1, H_2, \ldots, H_k$ be k mutually exclusive and exhaustive statements (events) and $P(H_i)$ their (prior) probabilities. Further, let

E be an observable event, which has alternative probabilities relative to the statements, $P(E|H_i)$. Then—after E is observed, the (posterior) conditional probability of the statements H_i is given by

$$P(H_i|E) = \frac{P(H_i) \cdot P(E|H_i)}{\sum_{j=1}^{k} P(H_j) \cdot P(E|H_j)}.$$

To understand why this has been called the inverse problem, we go back to the special case that Bayes solved. The direct problem thereby relates to the development of the relative frequencies with increased number of trials

> [...] where the Causes of the Happening of an Event bear a fixed Ratio to those of its Failure, the Happenings must bear nearly the same Ratio to the Failures, if the Number of Trials be sufficient; and that the last Ratio approaches to the first indefinitely, as the number of Trials increases. (Hartley 1749, cited from Hald 2007, p. 25)

The indirect problem now relates to the development of the "causes" in the light of observed relative frequencies (cause might be better read as relative ease or probability and has nothing to do with cause in the original sense).

> An ingenious Friend [it is agreed widely that Hartley refers to Bayes] has communicated to me a Solution of the inverse Problem, [...] where the Number of Trials is very great, the Deviation must be inconsiderable: Which shews that we may hope to determine the Proportions, and, by degrees, the whole Nature, of unknown Causes, by a sufficient Observation of their Effects. (Hartley 1749, cited from Hald 2007, p. 25)

Bayes used a material situation to embody his model, in which his assumptions clearly make sense. His model includes a parameter p for an unknown probability and a series of data simulated on the device that follow a binomial distribution with parameters n and p. From these conditions, Bayes derived an integral term for the probability of p to lie within the limits of p_1 and p_2. In fact, he derived the posterior distribution of p after the event E of k successes in n trials of the Bernoulli experiment is observed (which is a beta distribution). His "prior" distribution on the parameter p as embodied in his material model was a uniform distribution. Independently of this situation, Bayes gave another *objective* reasoning for his uniform prior (see also Edwards 1978). If you do not know anything about the combined experiment (including the parameter p), then the stakes for any number k of successes in n trials should be the same. From this assumption it is relatively easy to justify a uniform prior for the unknown parameter p.

It is interesting to note that the main purpose of Bayes to clarify the relation between probabilities (p) and relative frequencies of an experiment performed under the same conditions was forgotten in the later discussion on the status of this prior: the form of the prior distribution loses its influence on the limiting behaviour as the posterior distribution concentrates more and more around the—previously unknown—point p and converges to it. This is the inverse side of Bernoulli's law of large numbers.

One of the mathematical conclusions from Bayes' formula was the rule of succession as discussed by Price (1763–1764, 1764–1765) in his introduction to Bayes' article. If one observes an event k times repeatedly in n successive observations, its probability equals $(k + 1)/(n + 2)$; this converges to 1 for $k = n$ as n tends to infinity. From the many times, the sun rose already, Buffon derives a moral certainty the sun will arise every (future) morning in his *Essai d'arithmétique morale* (1777).

As Bayes gave objective reasons for his uniform prior, it was justified as the prior as if it were inherent in reality and not a model about it. That clearly gave rise to speculations. On the one side, it was claimed in the argument that one does not know anything about the (combined) experiment and also not about the parameter p. On the other hand, a uniform distribution was derived for it. Thus, it seemed inconsistent from a philosophical point of view. How can complete ignorance on something lead to a concept that yields obviously some information about the parameter? To enhance the dilemma, we refer to the special case of $n = 0$ and $k = 0$ when there is neither data nor knowledge about any outcomes of trials, and the probability for the event occurring equals $1/2$ by the rule of succession.

The root of this trouble is the basic misunderstanding between objectivist and subjectivist concepts. Of course, with enough (independent) data the influence of the prior vanishes leaving no difference in the solution between frequentist and Bayesian (subjectivist) statisticians. If there is not enough data, then the question of using priors arises. Without a prior there is no solution at all; the alternative is to use a prior to get a solution. However, the prior used should really be based on an expert knowledge. Yet, it will never reach a status of objectivity.

4 Foundations and New Applications

4.1 Classical Probability

Laplace marks a culmination in the early conceptual development of probability but he already stands at the edge of a new era, which is marked by extensive use of continuous distributions either as an approximation to the binomial or independently as describing phenomena in physics and "social physics". Yet, philosophically his views are still based on a mechanical determinism and linked simply to ignorance. The Laplacean demon has become notorious:

> Given [...] an intelligence which would comprehend all [...] for it, nothing would be uncertain and the future, as the past, would be present to its eyes [...] Probability is relative, in part to this ignorance, in part to our knowledge. (1812/1951, p. 4, 6)

Laplace gave what is the first explicit definition of probability, the so-called classical probability, which underlies the classical a priori theory (APT). The probability $P(A)$ of an event A equals the ratio of the number of all outcomes which are favourable to event A to the number of all possible outcomes of the trial. This definition implicitly assumes that the individual outcomes are equally likely. Laplace varied the (correct but not well understood) justification for the prior distribution in the Bayes' formula into a general 'principle of insufficient reason' to underpin this assumption. According to this principle, we may assume outcomes are equally likely if we have no reason to believe that one or another outcome is more likely to arise. It is easy to see how this may arise out of gambling situations where the assumption would be seen as so obvious that it would not even be seen as an assumption, rather like Euclid's axioms.

This first formal definition had a great impact on the prestige of probability and boosted its applications and the development of techniques (e.g., generating functions, see Laplace 1814/1995) but did not clarify the nature of probability, as it refers to a philosophically obscure principle for its application and as its domain is far from the actual applications (continuous distributions) being studied at that time. Ensuing attempts to amend this principle have tried to replace this by indifference or invariance considerations with the aim to lend the equal probabilities an objective character under restricted conditions. From a modern viewpoint, Barnett (1973, p. 71) states that the principle "cannot stand up to critical scrutiny". There were also some problems and ideas which hindered conceptual progress.

Independence was a steady source of difficulty as it was not dealt with as a distinct concept. In a notorious example of a famous mathematician erring in probability, d'Alembert (1754) opposed the equi-probability of the four outcomes in tossing two coins; he argued fiercely in favour of the probabilities $1/3$ for each of the outcomes 0, 1, and 2 'heads'. While he used a primitive argument on three possible cases, his solution can be restored by using the succession rule and prior ignorance. Equally his solution can be refuted by reference to the independence concept. It is remarkable that writers always blamed d'Alembert for a wrong equi-probability argument but did not report on the dilemma between the subjectivist and objectivist interpretations of the problem, which lead to differing solutions.

Conversely, independence was used implicitly for combining arguments, which were not at all independent as in calculating the probability of the judgement of tribunals to be correct: the error probabilities of single judges were multiplied as if they were independent. If the tribunal consisted of 7 judges and each of them erred with probability 0.03, then the probability of the majority being right was (incorrectly) calculated to be 0.99997, which was judged as morally certain.

If Laplace's principle of insufficient reason is taken for granted, it would yield another ambiguous rule to transfer ignorance into knowledge. Thus, the focus in using it has to be on finding a representation of a problem, in which the possible cases may be viewed as equally likely. However, even physical symmetries can give rise to different embodiments, which leave the problem of attributing equal probabilities to these cases inconclusive. As a model, this induces no problem, as the basis for a mathematical theory, it does.

Indeed, Fine (1973, p. 167) illustrates the difficulties by the example of three models for two dice. While the first model is generally seen as the standard, the other models in fact are useful to describe phenomena in physics and the real world (or at least theories about it).

Maxwell–Boltzmann model: each of the 36 pairs (1, 1), ..., (1, 6), (2, 1), ..., (2, 6), ..., (5, 1), ..., (5, 6), (6, 1), ..., (6, 6), is equally likely; pairs like (2, 3) and (3, 2) are different outcomes in this model.

Bose–Einstein model: each of 21 pairs (1, 1), ..., (1, 6), (2, 2), ..., (2, 6), ..., (5, 5), (5, 6), (6, 6), is equally likely; pairs like (2, 3) and (3, 2) are treated as identical outcomes.

> *Fermi–Dirac model*: each of 15 pairs (1, 2), ..., (1, 6), (2, 3), ..., (2, 6), ..., (5, 6), is equally likely; the two components are forbidden to have the same value.

For ordinary dice, the Maxwell–Boltzmann approach is the usual and apparently natural model. The two dice can be (at least theoretically) discriminated; blue and red dice, or first and second trial; the independence assumption is accepted as highly plausible. This supposedly natural model, however, does not fit some applications in physics. According to Feller (1957, p. 39), numerous attempts have been made

> to prove that physical particles behave in accordance with Maxwell–Boltzmann statistics but modern theory has shown beyond doubt that this statistics does not apply to any known particles.

The behaviour of photons, nuclei and atoms containing an even number of elementary particles which are essentially indistinguishable may be best described by the Bose–Einstein statistics. Electrons, protons and neutrons where the particles are essentially indistinguishable and each state can only contain a single particle are well described by the Fermi–Dirac model. It is quite startling that the world is found to work in this way, experimentally. The situation is comparable to cases where non-Euclidean geometry is actually found to be a better model of the real world than the "natural" Euclidean geometry.

Symmetries are dependent on the purpose of modelling. It is salutary and interesting to note that pedagogical problems also arise over this issue. Therefore, probability should not be viewed as an inherent feature of real objects but only as a conceptual outcome of our endeavour to model reality. The intention to derive unique probabilities leads to various paradoxes. Fine (1973, p. 167) gives many counterexamples and summarizes his critique against Laplace's definition:

> We cannot extract information (a probability distribution) from ignorance; the principle is ambiguous and applications often result in inconsistencies; and the classical approach to probability is neither an objective nor an empirical theory.

4.2 Continuous Distributions

Continuous distributions are used for the first time by de Moivre (1738/1967). He investigates the probability of deviations of a binomial variable H_n from its expected value np:

$$P\big(|H_n - np| \leq d\big)$$

He derives an approximation in terms of what we now call the normal distribution. For de Moivre the limiting expression served only as an approximation. Simpson (1755) in his attempt to give an argument for using the mean of observations instead of single observations used a triangular distribution. To describe error distributions, rectangular, triangular, quadratic distributions, and Laplace's double exponential distribution were used.

With his central limit theorem, Laplace (1810) paved the way for the normal distribution: the binomial distribution approaches the normal as the number of trials tends to infinity. Laplace immediately recognized its importance and formulated an intuitive rule when a normal distribution was to be expected: in analogy to the situation in the binomial case, which is built up by single summands of 1 or 0 (depending on the outcome of the like binomial trial), a variable should follow the normal law of probabilities if it is (virtually or really) built up by adding a large number of quantities. Following this interpretation, the normal distribution was soon the main distribution used in describing observational error or physical quantities.

There were two urgent problems arising from the more and more experimentally oriented sciences. The first was to give a probabilistic justification of the use of the method of least squares to derive an estimate for an unknown quantity from observations. The second was to give a probability argument for using the arithmetic mean instead of single observations, in order to "eliminate" measurement errors. Gauss (1809) uses the normal distribution in its own right and not only as an approximation within the context of error theory. His proof of the method of least squares was based on Laplace's inverse probability with a uniform prior on the location parameter and a normal distribution for the observations. Gauss also derived a functional equation from the condition that the mean is the best value—in the sense of the most probable value—to extract from a series of observations that lead to the normal distribution as solution. The concepts of maximum likelihood, method of least squares, and normal distribution are related and these interrelations supported their usage.

The new ideas were soon picked up by researchers, especially Laplace's intuitive argument was convincing: Bessel (1818) checked the validity of the normal distribution against empirical data of measurements. Quêtelet (1835) extended the idea of error distribution to biometric measurements developing the concept of l'homme moyen, an ideal figure that gets its value by the interference of a large sum of errors of nature. Thus, any biometric variable of human beings or animals should follow a normal distribution. The myth of the ubiquity of the normal distribution was born. Galton (1889) constructed the Quincunx to demonstrate how Nature would realize its elementary errors to add up to the final value of the variable under investigation. This apparatus is an elegant embodiment of the central limit theorem for the special case of a sum of binomial trials.

The enthusiasm with the normal distribution did not end with the systematic search of Pearson (1902) for systems of continuous distributions, nor with the Maxwell distribution for velocities in physics. Physics, especially the field of thermodynamics, saw a rapid development of concepts using the idea of distribution to describe the system's behaviour at the microscopic level in order to derive their laws of entropy on the macroscopic level. The developments in physics and especially in thermodynamics are seen as a driving force to develop a sound basis (see Batanero et al. 2005) and lead to Hilbert's (1900) programme to find an axiomatic basis for physics *and* probability. The more theoretical line of development was pursued by the Russian school (see Maistrov 1974): Following Poisson (1837), the Russians (Chebyshev 1887; Markov 1913) pursued various generalizations of the central limit theorem to derive the normal distribution of a sum under ever generalized conditions for the single parts of the sum.

4.3 Axioms of Probability

Various endeavours have been undertaken to develop a basis of the theory of probability following the idea of relative frequencies, culminating in the attempt by von Mises (1919) to give an axiomatic approach to the discipline based on idealized properties of random sequences, yet they were futile (Porter 1986) and amended only much later (Schnorr 1971). Since the turn of the century, the gold standard of a mathematical theory requires an axiomatic foundation. It is amazing that—despite Hilbert's (1900) agenda for the axiomatic foundation of physics and probability—it was not before 1933 that Kolmogorov was successful. This has to be contrasted with the great progress in measure theory around 1900. Lebesgue (1902) formulated the central problem of measure theory as the question of which sets can have a measure (as an extension of length, area, etc.). The problem was solved by the community using the class of Borel sets and the agreement (after some dispute) that the characteristic property of such measures should be captured by their sigma additivity. Borel (1909) applied this new knowledge to prove the strong law of large numbers and showed that the set of trajectories (of relative frequencies), which do not converge to the underlying probability has a zero measure. Such measures on sets of trajectories in three-dimensional space (which are elements of an infinite-dimensional space) have also become important to describe processes and laws in physics (especially in thermodynamics). What hindered the community to find a common logical base for this measure on the trajectories and the probabilities at the outset, which were still based on Laplace's equi-probability?

One obstacle surely was to find the common basis for discrete and continuous distributions. Another hindrance was the key concept of independence that is not contained in measure theory. Kolmogorov (1933), in his milestone publication, solved both obstacles. He came from almost innocent-looking axioms to the concept of distribution functions ($F_X(x) = P(X \leq x)$) with the result that there are certain types of distributions, including discrete and continuous distributions. For a succinct exposition of the mathematical ideas and concepts, see Borovcnik et al. (1991). Moreover, Kolmogorov dealt with independence as a key concept additional to the axioms. His approach is open to diverse interpretations of probability, including APT, FQT, and SJT, and has been taken as the mathematical foundation for probability.

Thus, on the one side, the prestige of the mathematical foundation let the applications boom again; on the other side, it laid the basis for a fierce controversy in the foundations starting with de Finetti (1937) who presented a system of axioms based on the interpretation of probability as degree of belief and ideal preference behaviour. Today, we can state that the battle remains unsettled and does not even clarify the various positions. However, the theories can be formulated in a way to conform the axioms of Kolmogorov. That is to say that most subjectivist followers of de Finetti would read the Kolmogorov axioms in the same way as objectivists but have their different interpretations. The schools would, however, differ considerably in their approach to statistical inference.

From the perspective of analytic science, Stegmüller (1973) agreed that statistical inference cannot be derived from the position of axiomatic probability alone but has to refer to further rationality criteria, which still have gaps. One unsolved problem is the difficulty to derive any direct statement for a single case from probabilistic information that is based on a long-run interpretation. There is no way to justify a significance level or power (of a test) or a confidence level (for confidence intervals) for the single application of a test. Popper's (1957) concept of propensity is just an elegant wording to the problem without a solution. According to that, probability is a physical property of an object or situation to produce events—a propensity.

From the view of applications, today the usage is liberal. If one has enough data then the use of methods of statistical inference without prior distributions on parameters leads to almost the same conclusions as subjectivist methods that make use of prior distributions, so there should be no dispute about the result even if the interpretation of it would still differ. If such frequentist information is lacking, the use of prior distributions has to be carefully calibrated against expert knowledge with the usual caveats in interpreting the final result as a guidance to decide but not as the final decision. What is difficult for probability is that the axiomatic foundation is not sufficient. The concepts cannot preserve their character once they are applied. For teaching, it means that the full comprehension of probability can neither be detached from an axiomatic perspective, nor from the inferential part of stochastics to preserve wider meaning.

5 Modern Philosophical Views on Probability

We now discuss the three main approaches to the nature of probability which are relevant for school mathematics. They are summarized below as classical a priori, frequentist, and subjectivist. The structural view is discussed as a sort of synthesis. We also present some virtues and some criticism of each approach; a fuller treatment can be found in Barnett (1973), or in Weatherford (1982), which also deals with other philosophical positions.

Classical a Priori Theory (APT) Following Laplace, the probability of a combined event is obtained by the fraction of outcomes favourable to this event in the sample space; this makes use of an implicit assumption of equal likelihood of all the single outcomes of the sample space. It is an a priori approach to probability in that it allows calculation of probabilities *before* any trial is made. Geometric probability is closely related to it; it reduces the probability concept to counting or area.

In the case of applications, one is confronted with the problem of deciding the single outcomes that are equally likely, leading to the charge of circularity. Symmetry in the physical experiment with respect to Laplace's 'principle of insufficient reason' is a shaky guideline to help in this respect. One major philosophical problem is that the same physical experiment can reveal several different symmetries (as for the dice above). Hence the criticism is that the logic of symmetry is useless for prediction. Another criticism is that APT is subjective with regards to choosing the underlying symmetry. Furthermore, it is not clear how APT should be adapted by experience—for example, in the case of a (biased) die which lands on 3 for ten consecutive tosses. Conversely, there is virtually universal agreement that the probability of a die (with no obvious bias) to land on 6 is $1/6$. There are children who deny this from their experience of board games, but this is linked to expectation (of waiting times till the first six) rather than to the probability of getting sixes. As we have seen, such examples from gambling were the source of the fundamental idea of probability. There was no intention that probability should apply to actuarial tables; for some it still seems odd to calculate the chance of dying at different ages, based on mortality tables. Thus some would say that there are two fundamental notions of probability: one is based on symmetries whilst the other is based on data.

Frequentist Theory (FQT) The probability of an event is obtained from the observed relative frequency of that event in repeated trials. Probabilities are never obtained exactly by this procedure but are estimated. It is an a posteriori, experimental approach based on information *after* actual trials have been done. The measure of uncertainty is assigned to an individual event by embedding it into a collective—an infinite class of 'similar' events which are assumed to have certain 'randomness' properties: these ideas were developed by von Mises. Then the probability is the limit towards which the relative frequency tends.

In applying this definition, one is faced with the problem that an individual event can be embedded in different collectives, with no guarantee of the same resulting limiting frequencies: one requires a procedure to justify the choice of a particular embedding sequence. Furthermore there are obvious difficulties in defining what is meant by 'similar' or by 'randomness'; indeed an element of circularity is involved. Even the notion of settling down presents difficulties in terms of the number of trials needed in long term frequency. More fundamentally, there are many events where a probability is required but it is not possible to carry out repeated experiments—this is especially true for events with a low probability. There are three main criticisms of FQT. The first is that FQT probabilities can, in principle, never be calculated. The second is that it cannot be known whether FQT probabilities actually exist. Finally, a tentative FQT value can never be confirmed or even denied. On the other hand, FQT probability is more suited to modern views and seems to be based on evidence rather than external thought. FQT also offers a practical means of calculating probabilities, especially in situations, such as mortality at different ages, where no symmetry can be applied. In some ways, FQT is seen as more scientific and certainly features in modern physics and biology.

Subjectivist Theory (SJT) Probabilities are evaluations of situations which are inherent in the individual person's mind—not features of the real world around us which is implicitly assumed in the first two approaches above. The basic assumption here is that individuals have their own probabilities which are derived from an implicit preference pattern between decisions. Ideas are taken from gambling where given any event, its probability can be determined by the odds a person is prepared to accept in betting on its occurrence. Obviously people may differ in the odds they would accept, but this does not matter provided basic rules of coherence and consistency are met. For example, it would be foolish to place two bets of 3 to 2 on both horses in a two-horse race (i.e. for a stake of £2 you could win £3) because one is bound to lose money as the win of £3 does not compensate for the loss of £4 overall. Coherence formalises this basic idea from which one can deduce the basic laws of probability (Barnett 1973). These ideas were developed in detail by de Finetti (1937), who starts his lifetime work (de Finetti 1974) by the startling claim that 'PROBABILITY DOES NOT EXIST'. His followers whilst being very vociferous, remain in a minority amongst mathematicians.

For a subjectivist there are two categories of information, namely prior information which is independent of any empirical data in a person's mind, and empirical data from frequencies in repeated experiments. Both types of information are combined in Bayes' formula to give a new probability of the event in question. This updating of probabilities is called 'learning from experience'. Bayes' formula, combined with simple criteria of rationality, allows a *direct* link from experience and prior information to decisions. Thus, the usual problems of statistical methods based on objectivist probability are circumvented.

A problem inherent to the subjectivist approach is its intended ubiquity; any uncertainty has to be specified by probabilities. There might be occasions when it is better not to force meagre information into the detailed form of a distribution; it might be wiser to try a completely different method to solve the problem. The most striking argument against the subjectivist position, however, is that it gives no guidance on how to measure prior probabilities (Barnett 1973, p. 227). Though there are flaws in the classical and frequentist approaches, they provide pragmatic and accepted procedures to calculate probabilities. Of course, a subjectivist could exploit all frequentist and symmetry information to end up with a suitable prior distribution; but there is nothing in the theory to encourage such an approach. Indeed any probability can be assigned to any event by an individual; the only constraint is to avoid incoherence such as described above for a two-horse race.

The initial prior probabilities can be changed as new information is obtained and this can be done using Bayes' theorem: this leads to the name of Bayesians for subjectivists, but Bayes' theorem is equally applicable in APT and FQT. The main criticism of SJT is its intrinsic subjectivity and lack of an objective basis: it seems to confuse feeling with fact. This implies that it does not matter what we believe and whether there is any evidence to the contrary, apart from the relatively less significant proviso to lack coherence. However, SJT is valid for cases where there is no symmetry for APT or evidence from repeated trials as required by FQT. For example, one may wish to discuss the probability that an incumbent president will win the next election. One may have a range of information about such an event and arrive at a probability; such an approach is certainly taken in the media. Indeed, it is often the case that such an approach is necessary in business decisions where knowledge is partial and yet a decision has to be made. One only has to think about the way that ideas of risk have become more prominent in the last decade, with the financial crisis and other events such as the spread of epidemics across the world. The SJT is ideally suited to such scenarios.

Commentary In the structural approach, typically taken in courses at higher levels, formal probability is a concept which is implicitly defined by a system of axioms and the body of definitions and theorems which may be derived from these axioms. Probabilities are derived from other probabilities according to mathematical theorems, with no justification for their numerical values in any case of application. This structural approach does not clarify the nature of probability itself though the theorems derived are an indicator of possible interpretations. The structural approach serves as a theoretical framework for the two main conceptions of probability.

The *objectivist* position encompasses the classical and the frequentist view; according to it, probability is a kind of disposition of certain physical systems which is indirectly related to empirical frequencies. This relation is confirmed by theorems like the law of large numbers. The *subjectivist* view treats probability as a degree of confidence in uncertain propositions (events). Axioms on rational betting behaviour like coherence and consistency provide rules for probabilities; the Kolmogorov axioms prove mathematical theorems which must be obeyed if one wants to deal rationally with probabilities.

Historically, physics was dominated by a causal approach, which also left its traces in the philosophy of randomness as may be seen by the Laplacean demon, which turned probability to the tool of those who are ignorant. More recently, however, this causal approach in physics has switched to a completely random approach where everything is referred back to random laws (see Styer 2000). Remarkably, sometimes the discussants forget about the use of a model in this approach and perhaps infer that the whole world is random. However, there are alternative approaches in the foundations of physics, which stress causality (e.g. Dürr et al. 2004). The causality—randomness debate will surely continue.

Within the connections between randomness and causality in physics, the controversy in the foundations of probability, between those who promote an objective character of probability (the Objectivists) and those who believe that probability is fundamentally subjective (the Bayesians), was fiercely fought but has been decided in favour of an interpretation of Kolmogorov's axiomatic probability as something like a propensity in the sense of Popper (1957). Such a decision was taken in analytic philosophy despite the gaps in rational argument of statistical inference based on probability. Stegmüller (1973) asserted that a radical subjective physics cannot be accepted and, to accept the gap in rationality, is less problematic. However, apart from physics, the Bayesian approach has been widely accepted in the sense of Berger (1993). If enough data is available, then a classical (objectivist) approach may be used, if not then qualitative information has to be used and the problem remains how to make it more communicable between those who use the model.

For teaching, the discussion in the American Statistician in 1997 has clearly shown some disadvantages of a classical approach for statistical inference (Albert 1997; Berry 1997; and Witmer et al. 1997). But as Moore (1997) states, Bayesian methods are too complicated to teach initially. In fact, Bayesian methods for statistical inference are easier to understand but—as they deal with qualitative data—much harder to apply, at least when frequentist information is missing. Vancsó (2009) devotes special attention to exploit the advantages of both approaches simultaneously in teaching.

This shows, on the one hand, the close connection of probability to physics, which will become even stronger as physics might abandon the causal paradigm in favour of a complete randomization ideal. On the other hand, statistical inference has evolved greatly by the sound foundation of probability in the last century, and revived the subjective roots of probability again. As it is not possible to close the gaps in rationality, the community follows the same pattern of reaction as in the case of Gödel's fundamental critique to mathematics, namely to use that approach which is more suitable in each situation without being able to universalize the foundations.

Finally, we turn to topical aspects of probability relating to risk. As may be seen from the historic development, risk has always played a prominent role in the development of the concepts. This is not restricted to games of chance but extends to more general situations under uncertainty. By risk usually we refer to an outcome, which has a high negative impact, which should be avoided. More neutrally, in some situations, some outcomes have a small impact or may have a benefit. The decision situation involves several options, all with different outcomes related to the impact (loss or win). Which of the decisions is best? For a perceptive discussion of situations involving risks in law or medicine, see Gigerenzer (2002).

The difficulties are increased by the fact that many such situations involve very small probabilities (e.g. the prevalence of *BSE*) but the potential impact is extremely negative. The other difficulty is to separate the awareness of the probabilities and their measurement from the perception of the impact. High (negative or positive) impact might get so dominant that the probabilities associated to it may be "forgotten". Moreover, as the probabilities are often so small, they cannot be measured by frequencies. This links to Bernoulli's discussion of how to combine stochastic arguments.

These topics pose a challenge, not only for teaching but also for further steps of conceptual development. For teaching, the challenge is discussed in Kapadia and Borovcnik (1991), Jones (2005), Borovcnik and Kapadia (2009), and Batanero et al. (2011). More recently, the perspective of modelling with multiple models in a problem situation has been explored by Borovcnik and Kapadia (2011).

6 Concluding Comments

The key results in probability theory are similar in both FQT and SJT, as well as in APT when it applies. At least this is true if one neglects the more radical subjectivist variants like de Finetti, who denies even the σ-additivity of probability measures. That means that the competing positions use the same axiomatic language with respect to syntax. Of course, the semantics are different; Kolmogorov's axioms are not the basic medium for the interpretation between theory and reality, but are logically derived rules. Despite the formal equivalence of the approaches, Kolmogorov's axioms are usually thought to be a justification of the objectivist, especially the frequentist view.

To illustrate the various philosophical positions, we present a standard example of tossing a particular blue die in which APT, FQT, and SJT can be used. The classical view would assign a probability of $1/6$ of getting a six as there are six faces which one can assume are all equally likely; the frequentist view would assign a probability by doing repeated experiments or, as experiments with other dice have yielded $1/6$, the same value would be assigned for the die in question. In both these viewpoints, the probability is seen as an inherent feature of the die, and the tossing procedure. The subjectivist would assert that the probability was a mental construct which he/she may alter if new information became available about the die. For example, a value different to $1/6$ might be assigned if the die is black or heavier or smaller or was slightly chipped. There ought to be some basis for the decision made, but this is not a requirement for de Finetti. However, since the probability is not viewed as inherent in the object, the conflicting values would not involve any logical problems. In fact, the subjectivist does not reject considerations of symmetry or frequency—both are important in evaluating probabilities. There is an expectation, however, that ideas of symmetry or frequency are made explicit if applied.

In any case of application, one has to choose a specific subjective or objective interpretation to determine the model and inherent probability. Philosophically, for the structuralist, there is no way of deciding which model is better. This controversy on the nature of probability will continue. Fine (1973) states that

> subjective probability holds the best position with respect to the value of probability conclusions, [...]. Unfortunately, the measurement problem in subjective probability is sizeable and conceivably insurmountable [...]. The conflict between human capabilities and the norms of subjective probability often makes the measurement of subjective probability very difficult. (p. 240)

This extensive historical and philosophical background provides a valuable perspective for teachers. It is for them to decide how to present ideas to pupils and students, whether by APT, FQT, or SJT. We assert that a judicious combination of all three approaches is required to build up firm concepts from emerging intuitions as has happened in history, leading to the fiercely competing philosophical positions.

References

Albert, J. (1997). Teaching Bayes' rule: a data-oriented approach. *The American Statistician*, *51*(3), 247–253.

Arbuthnot, J. (1710–1712). An argument for divine providence taken from the constant regularity observed in the birth of both sexes. *Philosophical Transactions of the Royal Society*, *27*, 186–190. Reprinted in M. Kendall, & R. L. Plackett (Eds.) (1977), *Studies in the history of statistics and probability* (Vol. 2, pp. 30–34). London: Griffin.

Barnett, V. (1973). *Comparative statistical inference*. New York: Wiley.

Batanero, C., Henry, M., & Parzysz, B. (2005). The nature of chance and probability. In A. G. Jones (Ed.), *Mathematics education library: Vol. 40. Exploring probability in school: challenges for teaching and learning* (pp. 15–37). New York: Springer.

Batanero, C., Burrill, G., & Reading, C. (Eds.) (2011). *Teaching statistics in school mathematics. Challenges for teaching and teacher education. A joint ICMI/IASE study: the 18th study*. New York: Springer.

Bayes, T. (1763). An essay towards solving a problem in the Doctrine of Chances. *Philosophical Transactions of the Royal Society*, *53*, 370–418. Reprinted in E. S. Pearson, & M. G. Kendall (1970), *Studies in the history of statistics and probability* (Vol. 1, pp. 131–154). London: Griffin.

Bellhouse, D. R. (2000). De Vetula: a medieval manuscript containing probability calculations. *International Statistical Review*, *68*(2), 123–136.

Berger, J. O. (1993). *Statistical decision theory and Bayesian analysis*. New York: Springer.

Bernoulli, D. (1738/1954). Specimen theoriae novae de mensura sortis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, *5*, 175–192. English translation: Sommer, L., Exposition of a new theory on the measurement of risk. *Econometrica*, *22*, 23–36.

Bernoulli, J. (1713/1987). *Ars conjectandi*. Basel: Impensis Thurnisiorum, Fratrun. Translation of 4th part by N. Meunier. Rouen: Institut de Recherche sur l'Enseignement Mathematique; German translation by R. Haussner (1899), *Wahrscheinlichkeitsrechnung*. Leipzig: Engelmann.

Berry, D. A. (1997). Teaching elementary Bayesian statistics with real applications in science. *The American Statistician*, *51*(3), 241–246.

Bessel, F. W. (1818). Untersuchungen über die Wahrscheinlichkeit der Beobachtungsfehler. *Astronomische Nachrichten*, *15*, 358–359, 368–404.

Borel, É. (1909). Les probabilités dénombrables et leurs applications arithmétique. *Rendiconti del Circolo Matematico di Palermo*, *27*(1), 247–271.

Borel, É. (1943). *Les probabilités et la vie*. Paris: Presses Universitaires de France.

Borovcnik, M. (2011). Strengthening the role of probability within statistics curricula. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics. Challenges for teaching and teacher education. A joint ICMI/IASE study: the 18th study* (pp. 71–83). New York: Springer.

Borovcnik, M., & Kapadia, R. (Eds.) (2009). Research and developments in probability education. *International Electronic Journal of Mathematics Education*, *4*(3), 111–130. Online: www.iejme.com/032009/main.htm.

Borovcnik, M., & Kapadia, R. (2011). Modelling in probability and statistics—key ideas and innovative examples. In J. Maaß, & J. O'Donoghue (Eds.), *Real-world problems for secondary school students—case studies* (pp. 1–44). Rotterdam: Sense Publishers.

Borovcnik, M., Bentz, H.-J., & Kapadia, R. (1991). A probabilistic perspective. In R. Kapadia, & M. Borovcnik (Eds.), *Mathematics education library: Vol. 12. Chance encounters* (pp. 27–71). Dordrecht: Kluwer Academic.

Buffon, G. L. (1777). Essai d'arithmetique morale. In G. L. Buffon, *Histoire naturelle générale et particulière* (Suppl. 4). Paris: Imprimérie Royale. English translation (1785), *Natural history, general and particular* (2nd ed.). London: Strahan and Cadell.

Cardano, G. (approx. 1564). Liber de ludo aleae, published 1663 in *Opera Omnia*, *1*. Lyon: Sponius.

Chebyshev, P. L. (1887). Sur deux théoremes relatifs aux probabilité. *Acta Arithmetica*, *XIV*, 805–815. French transl. 1890–1891.

Çinlar, E. (2011). *Probability and stochastics*. Berlin: Springer.

d'Alembert, J. (1754). Croie ou pile. In D. Diderot, & J. d'Alembert, *Encyclopédie ou Dictionnaire raisonné des sciences, des arts et des métiers* (Vol. 4, pp. 512–513). Paris: Société des Gens de Lettres.

Daston, L. (1988). *Classical probability in the enlightenment*. Princeton: Princeton University Press.

David, F. N. (1962). *Games, gods and gambling*. London: Griffin.

de Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Annales Institut Henri Poincaré*, *7*, 1–68. Foresight: its logical laws, its subjective sources (Chaps. 1–4 translated by H. E. Kyburg Jr.). In S. Kotz, & N. L. Johnson (1992), *Breakthroughs in statistics. Vol. I. Foundations and basic theory* (pp. 134–174). New York: Springer.

de Finetti, B. (1974). *Theory of probability*. New York: Wiley. Translated by A. Machi, & A. Smith.

de Moivre, A. (1738/1967). *The doctrine of chances* (2nd ed., fuller, clearer, and more correct than the first). London: Woodfall. Reprint of 3rd ed. 1967. New York: Chelsea. 1st ed. 1718, London: Pearson.

Dürr, D., Goldstein, S., Tumulka, R., & Zanghi, N. (2004, August, 24). Bohmian mechanics and quantum field theory. *Physical Review Letters*, *93*(9), 090402. Online: http://arXiv.org/abs/quant-ph/0303156.

Edwards, A. W. F. (1978). Commentary on the arguments of Thomas Bayes. *Scandinavian Journal of Statistics*, *5*, 116–118.

Feller, W. (1957). *An introduction to probability theory and its applications* (2nd ed.). New York: Wiley.

Fermat, P., & Pascal, B. (1679). Correspondence 1654. In P. Fermat: *Varia opera mathematica*. Toulouse: Joannem Pech. English translation in F. N. David (1962), *Games, gods and gambling* (pp. 229–253). London: Griffin.

Fine, T. L. (1973). *Theories of probability*. New York: Academic Press.

Galton, F. (1889). *Natural inheritance*. London: Macmillan.

Gauss, C. F. (1809). *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*. Hamburg: Perthes und Besser.

Gigerenzer, G. (2002). *Calculated risks: how to know when numbers deceive you*. New York: Simon & Schuster.

Gigerenzer, G. (2004). Die Evolution des statistischen Denkens. *Stochastik in der Schule*, *24*(2), 2–13.

Graunt, J. (1662). Natural and political observations upon the Bills of Mortality, chiefly with reference to the government, religion, trade, growth, air, diseases etc. of the City of London. *Royal Society of London*.

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

Hacking, I. (1990). *The taming of chance*. Cambridge: Cambridge University Press.

Hald, A. (1990). *A history of probability and statistics and their applications before 1750*. New York: Wiley.

Hald, A. (2007). *A history of parametric statistical inference from Bernoulli to Fisher, 1713–1935*. New York: Springer.

Hartley, D. (1749). *Observations on man, his frame, his duty, and his expectations*. London: Richardson. Reprinted 1996 with an introduction by T. L. Huguelet. Gainesville: Scholar's Fascimiles and Reprints.

Hilbert, D. (1900). Mathematische Behandlung der Axiome der Physik. *Akademie der Wissenschaften in Göttingen*, *1900*, 272–273. Online: http://aleph0.clarku.edu/~djoyce/hilbert/problems.html#prob6 (English translation). From the lecture "Mathematische probleme" delivered at the International Congress of Mathematicians in Paris 1900. Online: www.mathematik.uni-bielefeld.de/~kersten/hilbert/problems.html.

Huygens, C. (1657). De ratiociniis in ludo aleae. In F. v. Schooten: *Exercitationes matematicae*, Leyden: Elsevirii.

Jones, G. A. (2005). *Mathematics education library: Vol. 40. Exploring probability in schools. Challenges for teaching and learning*. New York: Springer.

Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, *XLVII*, 263–291.

Kapadia, R., & Borovcnik, M. (Eds.) (1991). *Mathematics education library: Vol. 12. Chance encounters*. Dordrecht: Kluwer Academic.

Kolmogorov, A. N. (1933). *Ergebnisse der Mathematik. 2. Band, Heft 3. Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin: Springer. Reprinted 1977 (62 pp.). English translation 1956: *Foundations of the theory of probability*. New York: Chelsea.

Laplace, P. S. (1810). Mémoire sur les approximations des formules qui sont fonctions de très grands nombres et sur leur application aux probabilités. *Mémoires de L'Académie des Sciences Paris*, *1809*, 353–415, 559–565.

Laplace, P. S. (1812/1951). Essai philosophique sur les probabilités. *Journal de l'École Polytechnique*, *VII/VIII*, 140–172. English translation: *A philosophical essay on probabilities* (extended version). New York: Dover.

Laplace, P. S. (1814/1995). *Théorie analytique des probabilités* (2nd ed.). Paris: Courcier. 1st ed. 1812. Reprinted 1995, Paris: Jacques Gabay.

Lebesgue, H. (1902). Intégrale, Longeur, Aire. Thése. *Annali di Matematica Pura ed Applicata*, *7*(3), 231–359. Reprinted in H. Lebesgue (1972/1973), Oeuvres scientifiques (Vol. 1, pp. 203–331). Genève: Université de Genève, Institut de Mathématiques.

Logos Bible Software (n.d.). *The Holy Bible*. English standard version. Containing old and new testaments. Online. http://biblia.com/.

Maistrov, L. E. (1974). *Probability theory. A historical sketch*. New York: Academic Press.

Markov, A. A. (1913). *A probabilistic limit theorem for the cases of academician A. M. Lyapunov*. Reprinted in A. A. Markov (1951), Selected works of A. A. Markov in the theory of numbers and the theory of probability (in Russian). Leningrad: Izdatel'stvo Akademii Nauk SSSR, 319–338.

Moore, D. S. (1997). Bayes for beginners? Some reasons to hesitate. *The American Statistician*, *51*(3), 254–261.

Ore, Ø. (1953). *Cardano: the gambling scholar. With a translation from Latin of Cardano's "Book on games of chance" by S. H. Gould*. Princeton: Princeton University Press. Reprinted 1965, New York: Dover.

Pacioli, L. (1494). *Summa de arithmetica, geometria, proportioni et proportionalitá*. Venedig: Paganino de' Paganini.

Pearson, K. (1902). On the systematic fitting of curves to observations and measurements—Part I. *Biometrika*, *1*, 265–303.

Peverone, G. F. (1558). *Due brevi e facili trattati, il primo d'arithmetica, l'altro di geometria*. Lione: G. di Tournes.

Poisson, S. D. (1837). *Recherches sur la probabilité des judgements en matiére criminelle et en matiére civile*. Paris: Bachelier.

Popper, K. R. (1957). The propensity interpretation of the calculus of probability and the quantum theory. In S. Körner (Ed.), *Observation and interpretation in the philosophy of physics*. London: Butterworth.

Porter, T. (1986). *The rise of statistical thinking, 1820–1900*. Princeton: Princeton University Press.

Price, R. (1763–1764). Introduction and appendix to Bayes' Essay. *Philosophical Transactions of the Royal Society of London*, *53*, 370–375; *54*, 404–418

Price, R. (1764–1765). A demonstration of the second rule in Bayes' Essay. *Philosophical Transactions of the Royal Society of London*, *54*, 296–297; *55*, 310–325.

Quêtelet, A. (1835). *Sur l'homme et le dévelopement de ses facultés, ou Essai de Physique Sociale*. Paris. Online: California Digital Library. www.archive.org/details/surlhommeetled00quet.

Schneider, I. (1989). *Die Entwicklung der Wahrscheinlichkeitstheorie von den Anfängen bis 1933—Einführung und Texte*. Berlin: Akademie-Verlag.

Schnorr, C. P. (1971). *Lecture notes in mathematics: Vol. 218. Zufälligkeit und Wahrscheinlichkeit. Eine algorithmische Begründung der Wahrscheinlichkeitstheorie*. Berlin: Springer.

Shafer, G. (1996). The significance of Jacob Bernoulli's Ars conjectandi for the philosophy of probability today. *Journal of Econometrics*, *75*, 15–32. Online: www.glennshafer.com/cv.html#articles (Vol. 55, preprint version), or www.sciencedirect.com/science/journal/03044076/75/1.

Simpson, T. (1755). A Letter to the President of the Royal Society, on the advantage of taking the mean of a number of observations in practical astronomy. *Philosophical Transactions of the Royal Society, Series A*, *49*, 82–93. Online: available at Google books.

Stegmüller, W. (1973). *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie* (Vol. 4). Berlin: Springer. First part: Personelle Wahrscheinlichkeit und Rationale Entscheidung. Second part: Personelle und statistische Wahrscheinlichkeit.

Stigler, S. M. (1986). *The history of statistics: the measurement of uncertainty before 1900*. Cambridge: Harvard University Press.

Styer, D. F. (2000). *The strange world of quantum mechanics*. Cambridge: Cambridge University Press.

Todhunter, I. (1865). *A history of the mathematical theory of probability from the time of Pascal to that of Laplace*. Cambridge and London: Macmillan & Co. Online: California Digital Library. www.archive.org/details/ofmathemahistory00todhrich.

Vancsó, Ö. (2009). Parallel discussion of classical and Bayesian ways as an introduction to statistical inference. *International Electronic Journal in Mathematics Education*, *4*(3), 291–322. Online: www.iejme.com/032009/main.htm.

Venn, J. (1888). *The logic of chance* (3rd ed.). London: Macmillan. 1st ed. published 1866.

von Mises, R. (1919). Grundlagen der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, *5*, 52–99.

von Plato, J. (1994). *Creating modern probability: its mathematics, physics and philosophy in historical perspective*. Cambridge: Cambridge University Press.

Weatherford, R. (1982). *Philosophical foundations of probability theory*. London: Routledge & Kegan Paul.

Witmer, J., Short, T. H., Lindley, D. V., Freedman, D. A., & Scheaffer, R. L. (1997). Teacher's corner. Discussion of papers by Berry, Albert, and Moore, with replies from the authors. *The American Statistician*, *51*(3), 262–274.

From Puzzles and Paradoxes to Concepts in Probability

Manfred Borovcnik and Ramesh Kapadia

Abstract This chapter focuses on how puzzles and paradoxes in probability developed into mathematical concepts. After an introduction to background ideas, we present each paradox, discuss why it is paradoxical, and give a normative solution as well as links to further ideas and teaching; a similar approach is taken to puzzles. After discussing the role of paradoxes, the paradoxes are grouped in topics: equal likelihood, expectation, relative frequencies, and personal probabilities. These cover the usual approaches of the a priori theory (APT), the frequentist theory (FQT), and the subjectivist theory (SJT). From our discussion it should become clear that a restriction to only one philosophical position towards probability—either objectivist or subjectivist—restricts understanding and fails to develop good applications. A section on the central mathematical ideas of probability is included to give an overview for educators to plan a coherent and consistent probability curriculum and conclusions are drawn.

1 How Paradoxes Highlight Conceptual Conflicts

What makes a paradox? Progress in the development of mathematical concepts is accompanied by controversies, ruptures, and new beginnings. The struggle for truth reveals interesting breaks highlighted by paradoxes that mark a situation, which reflects a contradiction to the current base of knowledge. Yet, there is an opportunity to renew the basis and proceed to wider concepts, which can embrace and dissolve the paradox. A puzzle, however, is a situation in which the current concept yields a solution that seems intuitively unacceptable. Such a puzzle shows that the intuitive basis of the concept has to be improved or that the concept is contrary to the expectation of the solver. Examples from other areas of mathematics include negative numbers (for younger children) or complex numbers (for most adults). From puzzles and paradoxes one can learn about crucial properties of the theory involved. The situations are challenging and can also lead experts to err; the purpose of the

M. Borovcnik (✉) · R. Kapadia
Institut für Statistik, Alpen-Adria Universität Klagenfurt, Universitätsstraße 65, 9020 Klagenfurt, Austria
e-mail: manfred.borovcnik@uni-klu.ac.at

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_3,

concepts can also be better understood than by a sequential exposition of theory and examples.

Székely (1986) is the standard exposition which classifies puzzles and paradoxes and highlights the crucial contradictions that have contributed to clarify the basis of probability and mathematical statistics. He discusses each paradox in five parts: history, its formulation, explanation, remarks, and references: the reader is encouraged to study his approach for an extensive range of puzzles and paradoxes, which include some presented in this chapter.

Our approach is to link more closely to the underlying ideas and their application in teaching and learning. This vision is directed at the multi-faceted concepts entrenched with philosophical principles, especially with the two grand schools and conceptions of probability which are linked to the frequentist and Bayesian (subjectivist) interpretation of probability. Our exposition also encompasses the personal thoughts of pupils and students who are encountering these ideas for the first time. The underlying ideas are complex, overlapping, and interwoven. Steps to explain and overcome the traps are needed. For learners, the concepts are still emerging and change their character. This affects the subject of probability, which itself affects other areas of study. Think, for example, of the drastic change of the paradigm in physics which currently has changed completely from a causal to a random ideal.

A major purpose is to explain what is paradoxical and to link the solution to mathematical concepts and their historical perception. We assert that the present dominant position of objectivist probability narrows the flexibility not only of the models but also of the conceptions of the learner with the consequence that the exclusion of subjectivist notions hinders understanding.

We discuss 15 paradoxes and puzzles, only five of these feature in Székely (1986). The paradoxes are grouped alongside mathematical concepts which sometimes deviate from the historical development. Equal likelihood is followed by a review of the principle of insufficient reason, which regulates how and when equally likely cases are present (Sect. 2). A discourse on expected values follows to cover the central competing idea to probability since olden times (Sect. 3). The frequentist conception of probability has emerged as almost the sole interpretation, despite some key puzzles on randomness (Sect. 4). The concept of conditional probability is connected to subjectivist interpretations of probability (Sect. 5). We present mathematical theory from school to university (Sect. 6), although in a curtailed format. While the paradoxes highlight isolated developments in concepts, this section is intended to reveal central ideas that *link* the concepts coherently. This requires more technical and mathematical detail, which sometimes is prone to be avoided. However, a deeper understanding of the fundamental ideas is vital and may get lost if one strives to simplify the mathematics too much. The final section (Sect. 7) presents our conclusions.

Our presentation shows that the mathematical context of the concepts has to be accompanied by philosophical aspects, otherwise a comprehension of the theory will be biased, resulting in difficulties not only to understand but—more importantly—for learners to *accept* the concepts and apply them sensibly.

2 Equal Likelihood

The classical a priori theory (APT) starts with an assumption of equal likelihood, often in games of chance, and uses combinatorial methods to find the probability of various events. It is still, rightly, the initial approach to introduce probability to children. These ideas originally arose from a variety of puzzles and paradoxes.

Combinatorial multiplicity is linked to the possible outcomes which are considered to be equally likely. Moreover, links are soon made to relative frequencies as otherwise the concept of probability would lack an orientation about what will happen in repeated experiments. The emergence of these ideas is interwoven. To investigate equally likely cases and to apply a rule of favourable to possible cases draws heavily on counting the possibilities correctly, which proves to be harder than one might imagine, as evident from the difficulties shown by children and students. The conceptual confusion was aggravated by a competing concept, the expected value: at times, some like Huygens (1657) regarded it as more basic than probability with his value of an enterprise, corresponding to today's expected value. This was shortly after the great leap forward by Pascal and Fermat in 1654 (Fermat and Pascal 1679) when they specified a suitable set of outcomes and counted the possibilities correctly. An explicit, if contested definition of probability had to wait till Laplace (1812/1951). His approach is characterized by an intermixture of sample space—the mathematical part—and the intuitive part of an idea of symmetry. Probability is defined by assuming the equal likelihood of all possible results (APT).

In modern theory, the sample space is separated from probability, which is a function defined on a specific class of subsets of the sample space. This gives the freedom to view any specific probability as a model for a real situation. Until this final step to separate the levels of the model and the real problem, probability was considered as a property of the real world like length or weight. Thus—within bounds of measurement errors—there is *one* probability, a unique value for a problem. Accordingly, the task of a probabilist was either to find a sample space suitable to fit the equally likely conditions, or, since the empiricism of the nineteenth century, to find a suitable random experiment, repeat it often enough, and substitute the unknown probability by the relative frequency (FQT). Laplace recognized the difficulty related to judging cases to be equally likely and modified a principle going back to Jakob Bernoulli, which was later re-evaluated by Bayes: If one is ignorant of the ways an experiment ends up and there is no reason to believe that one case will occur preferentially compared to another, the cases *are* equally likely. This *principle of insufficient reason* underpins the application of Laplace's probability.

The first subsection below includes the historically famous puzzles arising from the struggle to sharpen the conception of possible cases, the rule of favourable to unfavourable cases, and expected value. Subsequently, problems arise from the principle of Laplace about equally likely cases. These provide excellent and motivating starting points to introduce probability in the classroom.

2.1 Early Notions of Probability

When concepts emerge and are not yet well-defined, confusion between closely related terms occurs. This may be confirmed by the early endeavours to find tools to describe and solve problems with uncertainty. As no other embodiments of the pre-concepts were available, it is no wonder that the context of games of chance was used extensively. Furthermore, the old idea of fairness and its close connection to games of chance was used as a "model" to a situation that had no relation to chance before.

The first puzzle, 9 or 10, deals with the sum of three dice. It marks a definite step towards the Cartesian product for counting the possibilities of repeated experiments, thus viewing the result 1.2.6 as different from 2.1.6, which gives a total multiplicity of 6 instead of 1. Lacking a theoretical argument for respecting order, the choice was justified by empirical frequencies. De Méré's problem with sixes marks a step in clarifying what can count as possibility. The confusion about the rule involved might be traced back to an overlap between the concepts of probability and expected value. The third puzzle is remarkable insofar as a counter-intuitive step was needed to sharpen the concept of possible cases: the solution is based on *hypothetical* cases, which are extended against the rules and are thus *impossible*. However, the greatest progress by Pascal and Fermat was to model a situation without a link to probabilities by a hypothetical game of chance to mimic the progress of a competition and use the resulting relative winning probabilities for a fair division of the stakes.

***P*₁: Problem of the Grand Duke of Tuscany** Three dice are thrown. The possibilities to get a sum of 9 or 10 are counted in the following way (Galilei 1613–1623; cited from David 1962, p. 192):

> [...] 9 and 10 can be made up by an equal diversity of numbers (and this is also true of 12 and 11): since 9 is made up of 1.2.6, 1.3.5, 1.4.4, 2.2.5, 2.3.4, 3.3.3, which are six triple numbers, and 10 of 1.3.6, 1.4.5, 2.2.6, 2.3.5, 2.4.4, 3.3.4, and in no other ways, and these also are six combinations.

This theoretical argument is confronted with experience:

> Nevertheless, although 9 and 12 can be made up in as many ways as 10 and 11 respectively, and therefore they should be considered as being of equal utility to these, yet it is known that long observation has made dice-players consider 10 and 11 to be more advantageous than 9 and 12.

Galilei ordered the results of the ways of getting 9 as follows: there are 6 different orderings of 1.2.6 but only 3 out of 2.2.5 and only one from 3.3.3. His table is well worth reproducing and studying by pupils (see Table 1), as its structure conveys a hierarchical process of ordering, first the results of the dice and then ordering them using the symmetry that a sum of 3 and 18 has the same multiplicity, as have 4 and 17, up to 10 and 11.

What is the Paradox? Two counting procedures lead to different numbers and yield different probabilities. As the probabilities were communicated in odds, the

Table 1 Galilei's protocol (from David 1962, p. 194)—slightly modified

10		9		8		7		6		5		4		3		
6.3.1	6	6.2.1	6	6.1.1	3	5.1.1	3	4.1.1	3	3.1.1	3	2.1.1	3	1.1.1	1	
6.2.2	3	5.3.1	6	5.2.1	6	4.2.1	6	3.2.1	6	2.2.1	3					
5.4.1	6	5.2.2	3	4.3.1	6	3.3.1	3	2.2.2	1							
5.3.2	6	4.4.1	3	4.2.2	3	3.2.2	3									
4.4.2	3	4.3.2	6	3.3.2	3											
4.3.3	3	3.3.3	1													
	27		25		21		15		10		6		3		1	108
																108
																216

first count leads to 1: 1 while the second yields 25: 27 for 9 against 10, which is advocated as correct by Galilei; today, we read the result as a probability of 0.1157 for 9 and of 0.1250 for 10. With the modern concept of repeated experiments, *independence* is a theoretical argument in favour of the second way to count. Without such a concept, the result was certainly puzzling. Another strange feature is how they could find such a small difference by playing. For Székely (1986, p. 3) the paradoxical feature of 9 or 10 lies in the fact that for two dice 9 is more probable while for three dice 10 is more probable.

Further Ideas The argument to find which solution is right is interesting. It signifies that since olden times counting the possibilities in order to calculate the relative probabilities was linked to what actually happens in games. As the difference in probabilities for 9 and 10 is very small (0.0093), it is hard to believe that this has really been detected by playing as this would require roughly 10,000 trials.

Such an argument as a substitute for theoretical reasoning is used at several places in the history of probability. It signifies that the theoretical argument alone was too weak to convince and that the writers considered a strong connection from their chances to relative frequencies. If an argument for a way to count contradicts the experience of relative frequencies, it is useless.

From a teaching perspective, there are valuable lessons which can be drawn and used. It is difficult for children to find all the possibilities in throwing three dice and then calculate the probabilities. Nevertheless, the approach here can help to develop combinatorial skills. It also confronts the difference between a theoretical (APT) probability and a frequentist (FQT) interpretation.

P_2: De Méré's Problem In this famous problem, a simple proportional argument suggests that it is equally likely to get (at least) a six in throwing a die four times as to get (at least) a double six in throwing two dice in 24 trials. De Méré posed the problem to Fermat as to why, apocryphally, he won a fortune betting on a six with one die and lost it betting on a double six in 24 trials (cf. David 1962, p. 235). Fermat listed the cases correctly and calculated the winning probabilities as $1 - (5/6)^4 = 671/1296 = 0.518$ for the six and as $1 - (35/36)^{24} = 0.491$ for the double six game. He concluded that, in fact, to bet on the single six game is favourable (higher than $1/2$) and betting on the double six game is unfavourable (lower than $1/2$).

What is the Paradox? In evaluating the chances, the following rule was emerging but far from clear: compare the number of favourable to unfavourable cases, or determine the ratio of favourable to possible cases. In a careless application of the emergent rule of "favourable to possible", the argument might have been as follows: with one die, 4 throws make 4 chances (i.e. 4 favourable cases to get a six). The 6 faces of the die mark the 6 possible cases. The ratio of favourable to possible yields $4/6$. With two dice, the 24 throws establish 24 chances (favourable cases to get a double six) with 36 possible cases and the same rule yields $24/36$. As the ratios are equal, the probabilities should be equal.

This argument is confronted by data in actually playing, whereby the game with one die is favourable while with two dice it is unfavourable. The line between favourable and unfavourable was drawn by the winning probability of $1/2$. If probabilities are linked to relative frequencies then there is definitely a problem with the original solution, and it is difficult to see what is wrong with counting the cases.

Another paradoxical feature lies in the difference between the concepts of probability and expected value, which are often confused in the discussion. The expected number of sixes in four trials with one die equals to $4 \cdot \frac{1}{6} = \frac{4}{6}$ and for double sixes with two dice it equals $24 \cdot \frac{1}{36} = \frac{24}{36}$. In this respect, a correct application of expected value leads to the same result for both games and the question is why this fails to predict the relative frequencies in games.

Further Ideas The classical random experiment is an independent binary 0–1 experiment with probability of p for the result 1. The expected value for one trial is p, for n repetitions it is $n \cdot p$. The reader may note that this yields another rule of favourable (the favourable cases are the n trials) to possible cases (with equal chances $1/p$ is the same as the number of all possible cases):

Expected value $= n \cdot p = n \cdot \frac{1}{1/p}$, which equals the fraction of n chances *to* $(1/p)$ possible cases.

Though the rules are identical they bear a different meaning, which can be recognized only if the difference between the concepts of probability and expectation is discussed in teaching.

This overlap may also be traced in Huygens' method to derive probabilities by calculating the corresponding expected values. However, in de Méré's problem, the two solutions differ. With the games above, the amount to win is 1 if the winning figure occurs (and 0 else); note that the payment is the same irrespective of the actual count and so the double six game is unfavourable. If the amount to win were exactly the *number* of sixes (double sixes), the games, in fact, would be equal. Yet, the double six game has a greater variance and bears more risk to lose but also gives more chances to win a higher amount.

Historically it was difficult to separate the concepts of probability and expectation and arrive at a clear vision of what probability can achieve and how to interpret or evaluate specific probabilities. The situation is usually blurred by the fact that an outcome is related to an impact, especially in games of chance. The perceived impact differs if it is a win or loss—even if the expected amounts are the same, as experiments by Kahneman and Tversky (1979) have shown.

P_3: Division of Stakes

> *A* and *B* are playing a fair game of balla. They agree to continue until one has won six rounds. The game actually stops when *A* has won four and *B* three. How should the stakes be divided?

Pacioli suggested the stakes should be split as 4 to 3 (Pacioli 1494; see David 1962, p. 37). We changed his original data to the situation Pascal and Fermat deal with in their famous exchange of letters 1654. They assess the possible ways to win the whole series:

> Since *A* needs 2 points and *B* needs 3 points the game will be decided in a maximum of four throws. The possibilities are: [see Table 2]. In this enumeration, every case where *A* has 2, 3, or 4 successes is a case favourable to *A*, and every case where *B* has 3 or 4 [successes] represents a case favourable to *B*. There are 11 for *A* and 5 for *B*, so that the odds are 11: 5 in favour of *A*. (Pascal in Fermat and Pascal 1679, referenced by David 1962, p. 91)

What is the Paradox? The solution is paradoxical from the standpoint of counting the *possibilities*, rather than dividing the stake by the current score of 4:3. What may be viewed as a possibility as the game has been interrupted? A great step forward is marked by introducing a hypothetical continuation of the game on the basis of what could happen if the game is continued. Of course, the series is decided if *A* wins two more games in this scenario. Thus, there are only 10 actual possibilities and 6 are favourable to *A*, which would split the stakes as 6 to 4 or 3 to 2. At this point, it is important that Fermat recognized that it makes no sense to consider these "real" cases as equally likely. To assign equal weights to them, he extended the "real" cases by imagined further rounds to *make* them of equal length. Interestingly, this conflicts with the rules of the game as one of the players would already have won and the series finished. To introduce a hypothetical continuation was the first step

Table 2 Pascal's hypothetical cases and real cases compared

[Counting hypothetical cases by Pascal					Counting "real" cases]				
AAAA	*AAAB*	*AABB*	*ABBB*	*BBBB*	*AA*	*ABA*	*BAA*	*BBAA*	*BBB*
	AABA	*ABAB*	*BABB*			*ABBA*	*BABA*	*BBAB*	
	ABAA	*BAAB*	*BBAB*			*ABBB*	*BABB*		
	BAAA	*ABBA*	*BBBA*						
		BABA							
		BBAA							

to *find* possibilities, to introduce an extension beyond the rules was to *invent* cases, with the aim to make them equally likely.

There are no signs that the stakes puzzle has anything to do with chance. The basic paradox is, however, Pascal and Fermat's view of the situation *as if* it were random. Originally the problem seems to be devoid of probability. The series has been interrupted and it is a problem awaiting a resolution. Probabilities are introduced only in the sense of a scenario. The focus lies on dividing the stakes fairly instead of finding a good model to describe the continuation of the game.

Further Ideas Székely (1986, p. 10) exclaims that "this [...] is considered [...] to be the birth of probability theory [...]". In earlier times, (fair) decisions were made by a game of chance if sufficient knowledge about the situation, or expertise (or trust) was missing. Somehow, probability introduces a sort of higher capacity (knowledge) beyond god (according to Laplace) who can always predict the result. Instead of exploring something like god's decision by a chancy game, the probability model is utilized to advocate a split of the stakes as fair. For teaching purposes, such a theoretical extension requires careful discussion in the classroom; it is instructive to explore the difference between a current proportion (4:3 here) and a fair division (11:5).

2.2 Conceptual Developments in Probability

The use of probability in the eighteenth and nineteenth centuries is signified by a diversity of conceptions. Two fundamental theorems concerned the relation between relative frequencies and probabilities: Bernoulli's law of large numbers (1713/1987) and Bayes' theorem including a corollary (1763).

Bernoulli's *direct probability* approach used the unknown probability p as a constant and the binomial distribution to derive the convergence of the relative frequencies towards this probability. With his theorem, Bernoulli provided the basis for relative frequencies (FQT) as an input to evaluate the probability of arguments. However, he was well aware that one might need more data to reach a reliable result.

Bayes' *inverse probability* method took the opposite approach: the weights on the unknown p converge to the relative frequencies of repeated experiments. Here, the unknown probability p is different from a constant (but unknown) number: in fact, one has to express a distribution upon it, which represents the status of knowledge on this parameter.

Bayes (1763) derived his theorem within an embodiment, in which the uniform prior distribution on p was obvious. Furthermore, he argued that if one lacks any knowledge about the value of the probability p of an event then one should accept equal stakes in betting on $0, 1, \ldots, n$ events in n repeated trials. On this assumption, he *derived* the uniform distribution on p *mathematically*. His argument was quite complex so that following writers abbreviated it to "if one lacks any knowledge about p then it is uniformly distributed" (similar to an older argument by

Bernoulli), which has become known as Bayes' postulate. Bayes' rule of succession is a corollary to his theorem and proves—on the basis of a uniform distribution on the parameter—that the posterior weights (distribution) on p after k successes in n trials have an expected value of $(k+1)/(n+2)$ and a variance that converges to zero (in modern terms he derived a beta distribution for the parameter p). Thus, the posterior distribution restricts itself to a point, which corresponds to the limit of the relative frequencies. Instead of using variance, Bayes calculated the probability of intervals around $(k+1)/(n+2)$, which was correctly interpreted as the probability of an event occurring in the next trial to follow (Price 1764–1765).

De Moivre gave a specification of probability as the number of favourable divided by the number of possible cases based on equally likely cases, which comes close to the generally accepted definition by Laplace (1812/1951). Laplace reproduced Bayes' theorem and used his "postulate" extensively to check and justify whether equally likely cases are appropriate in a problem. His approach ("if we are equally undecided about") was later named as the principle of insufficient reason. This was rejected by the empirical critique of Venn (1888). There is no way to transform complete ignorance into probabilities, which represent a form of knowledge. Venn asked for an *empirical* basis of probability as an idealized relative frequency. The difficulty of the principle of insufficient reason, and of independence is highlighted by the following problem where SJT is contrasted to APT.

***P*₄: D'Alembert's Problem** Two coins are flipped. What is the probability to obtain heads twice?

(a) Applying Laplace's principle on the combinatorial product space of *HH*, *HT*, *TH*, *TT* (thus respecting order) yields the answer $1/4$ (APT).
(b) D'Alembert (1754) refers to the fundamental probability set {*no head*, *one head*, *two heads*} (neglecting order) and applied equi-probability to the three cases, giving an answer of $1/3$ (will be linked to SJT below).

Since Pascal and Fermat it had been well acknowledged that, for repeated experiments, respecting order helps to find equally likely cases. Therefore, d'Alembert's approach was rejected as mere error illustrating how experts can err with probability (see Székely 1986, p. 3, or Maistrov 1974, p. 123).

What is the Paradox? The problem is a paradox in the history of probability as it has been overlooked that d'Alembert's solution *is* correct if only the *same* principle of insufficient reason is applied. This principle uses the uniform prior distribution on the unknown probability p of the coin to land heads up. With Bayes' rule of succession (a correct mathematical theorem) and the multiplication rule (a correct theorem) the following holds.

Before any data ($k=0$, $n=0$), one can conclude that $P(H)=\frac{0+1}{0+2}=\frac{1}{2}$; after seeing $H(k=1,n=1)$, the "conditional" probability to see heads again is $P(H|H)=\frac{1+1}{1+2}=\frac{2}{3}$. This yields the following probabilities for the three basic cases:

$$P(\textit{two heads})=P(H)\cdot P(H|H)=\frac{1}{2}\cdot\frac{2}{3}=\frac{1}{3},$$

$$P(no\ head) = P(T) \cdot P(T|T) = \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3},$$

and for the last it has to be

$$P(head\ tail\ mixed) = 1 - \frac{2}{3} = \frac{1}{3}.$$

The paradox leads to two probabilities (one linked to classical and the other to Bayes' conception of probability), which are irreconcilable.

Further Ideas There can be a strong justification to neglect the order of the results. D'Alembert's argument is theoretically consistent but neglects long-term experimental data on coins, which strongly supports the independence of the throws and corroborates the equi-probability of the four cases with the order distinguished. For a biased coin with probability p for head, the result would be p^2, $p(1-p)$, $(1-p)p$, $(1-p)^2$.

The underlying ideas are certainly worthy of careful discussion in the classroom as, in modern physics, the description of various particles in real physical systems can be described by d'Alembert's approach: photons, nuclei and atoms containing an even number of elementary particles are essentially indistinguishable and may be best described by the Bose–Einstein statistics, which *neglects* order. It is quite startling that the world is found to work in this way, experimentally.

3 Expectation

Historically, the concepts of probability and expected value developed in parallel with overlap and confusion. The first formalization by Huygens (1657) used the expected value in a financial framework based on a situation of implicitly equally likely cases. While the value of an enterprise is unambiguous, probability has been laden with personal conceptions and philosophical difficulties. Of course, both concepts are closely connected and Huygens used this economic value to calculate probabilities in problems, which were discussed at the time. The shift to probabilities was completed by Jakob Bernoulli (1713/1987) with his efficient combinatorial methods, which were faster than the rather lengthy recursive approach of Huygens.

Probability at the same time became strongly connected to relative frequencies by the law of large numbers by Bernoulli. As probability took the lead and expected value became a derived concept, the latter was engulfed by the philosophical "burden" of probability. Already in the publication of Huygens' treatise, the wording *expected value* appeared but this was mainly due to a bad translation to Latin and missed Huygens' intention. Such a change in terminology shifted away from the original frugal meaning of an economic exchange price between risk and certainty to wishes, desires, and similar vague conceptions.

Expected value lacked a strong connection to relative frequencies even if today it is motivated as the average amount paid after a long series of random experiments.

It is important to remember that expected value plays a basic role for the subjectivist position as probabilities get a wider interpretation, which integrates relative frequencies (if available) *and* qualitative knowledge beyond that. By such a weighting process, the subjectively accepted equivalence of a price (an expected value) in a simple 0–1 bet with probability p for 1 "measures" the personal probability of an individual.

3.1 Expectation and Probability

From a modern perspective, it has been forgotten that one may base the whole theory of probability either on the concept of probability or on the concept of expected value (despite the fact that there are mathematical approaches to reconstruct or replace Kolmogorov's axiomatic theory on the basis of expectation). It is no wonder that the two concepts were intertwined in the early stages. Huygens marks a special point in history—a temporary shift away from probability to expectation. He defined the term in an *economic* context as a price one would accept to switch between an uncertain (risky) situation and a situation without uncertainty—which resembles the features of taking out an insurance policy and amounts to a basic paradigm in decision theory.

The fundamental concept for Huygens is his value of an enterprise. Based on equal cases, he states

> to have p chances of obtaining a and q of obtaining b, chances being equal, is worth $\frac{pa+qb}{p+q}$. (Huygens 1657, cited from David 1962, p. 116)

Huygens circumvents the need to calculate probabilities or proportions; instead he solves the problems by his economic approach. The wording expected value is unfortunate as it associates hope, fear, and many other emotions related to the potential outcomes while Huygens used the term as a purely financial concept. In an analogy to determine the net present value of a future amount by a discount rate in financial mathematics, the present value of an uncertain enterprise equals the various amounts to gain or lose, *discounted* by their chances. Such a value is vital for any insurance policy. Future potential amounts have to be discounted to a value that is paid today.

The St Petersburg paradox unexpectedly (no pun intended) produced an infinite expected value, which is absurd if the concept is interpreted as an economic notion. The startling situation was resolved by amendments to probabilities, which are still disputed.

***P_5*: St Petersburg Paradox** Two players A and B toss a coin until it shows 'head' for the first time. If this occurs at the nth trial, then player B pays £ 2^{n-1} to player A. What amount should A pay to B before this game starts in order to make it fair? The expected value is infinite as the series diverges (cf. Székely 1986, p. 27):

$$2^0 \cdot \frac{1}{2} + 2^1 \cdot \left(\frac{1}{2}\right)^2 + 2^2 \cdot \left(\frac{1}{2}\right)^3 + \cdots = \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \cdots = \infty.$$

What is the Paradox? As an economic value of exchanging risk and certainty, an infinite value is completely unacceptable. No one could pay an infinite amount of money in advance. Nor would the player ever have a positive balance as all payments from the game will always be finite. And no human could witness the theoretical never-ending of tossing to get this infinite amount. A revision of the concept of expected value or of probability was urgently required. As an equivalent for an uncertain situation, expected value was regarded as a property of the situation. The time was not yet ripe to see this as artificial. However, it is still disputed whether a distribution makes sense if its expected value is infinite.

Further Ideas The paradox was put forward in 1738 by Daniel Bernoulli. Either the concept of expected value has to be revised or probability has to be conceptualized in a different way. The ways to resolve the paradox were twofold.

One way is to introduce *utilities* instead of money. Bernoulli (1738/1954) suggested replacing the payments by their logarithms arguing that the more money one has the less it is of importance to the person. In fact, he brought the expected utility down to a finite value but failed to provide a comprehensive solution to the paradox as, with a slightly different payment table, an infinite (expected) value would still result. The second way is to introduce a new entity such as a moral probability, which can be neglected if sufficiently small. The ensuing dilemma is to specify the size where probabilities lower than this benchmark could be neglected. The suggestions varied from 10^{-4} to 10^{-15}.

The concept of utility has been taken up by various approaches to applied probability. For a Bayesian probabilist, the way to evaluate an unknown probability is first by the subjective degree of credibility. Combinatorial multiplicity is a substantial factor as well as the information on past relative frequencies from similar experiments, but there are also personal and qualitative ingredients. All factors are prone to utility of the outcomes as—in measuring them—Bayesians would use the idea of equivalent bets that are accepted. Nowadays, utility has been revived by the discussion of teaching approaches based on risk which focus not only on probability but also on the impact associated to the possible outcomes.

The probabilist community has still not solved the problem of small probabilities. On the one hand, for events with small probabilities, the related impact may bias the personal perception of the magnitude, and data is missing as the probabilities are so small. On the other hand, small probabilities play a vital role as inherent properties of statistical procedures as the size of a significance test or the confidence level of a confidence interval show. Both reveal a lack of interpretation of small probabilities in the frequentist sense despite widespread endeavour to simulate the underlying assumptions in scenarios of the real situation.

3.2 Independence and Expectation

Probabilities have to be recalculated when games of chance are dependent. However, for expectation, it is irrelevant whether games are dependent or independent. In this sense, expected value is a functional analytic and not a stochastic property while variance is a genuine stochastic concept. In the first example below, two simple independent spinners are introduced and then changed to become dependent. In the second example, samples from a bag of coins are dealt with to illustrate the consequences of replacing or not replacing the drawn coins. The latter example highlights the advantages of the economic concept of a value.

***P_6*: Dependent Spinners** Two simple spinners are spun and the shaded area gives an amount of 1 to the player while the white sector leads to a 0 payment. The small spinner has a winning probability of p, i.e. $P(X = 1) = p$, and the big spinner of q, i.e. $P(Y = 1) = q$. The expected amounts to win are $E(X) = p$ and $E(Y) = q$. If played independently, one after the other, the fair price is p for the small and q for the big spinner (Fig. 1(a)). The price could also be paid in advance to play the game of $X + Y$ with an expected value of $E(X + Y) = p + q$. Putting the spinners one over the other, the game can be decided in one turn; when the spinner lands in the overlapping sector, the player wins both from the small and the big spinner, that means the win is 2 (Fig. 1(b)). With this variation the expected value of $X + Y$ remains the same. Whether the games are independent or dependent, for expected values the following additivity holds:

$$E(X + Y) = E(X) + E(Y).$$

What is the Paradox? Despite the close connection of expected value to probability, which remained confused for quite a long period, some basic properties differ. Expected values can be calculated from dependent random variables as if they were independent and represent—in this sense—*not a stochastic property*. Yet, the concept is used to find the fair price of a game of chance. In the special case of binary variables with 0 and 1, expected values do actually coincide with the probabilities.

Further Ideas With an overlap of x for the winning sectors in the spinners, the probabilities for the single payments are easily read off Table 3. The exact value of p_0 is not required for further calculations.

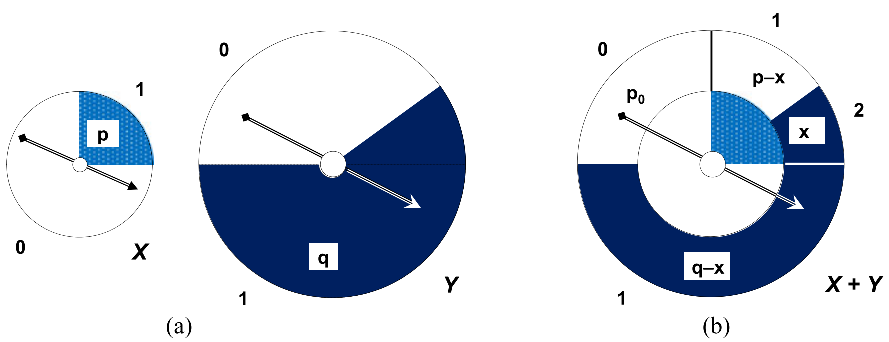


Fig. 1 (**a**) Two independent spinners. (**b**) Dependent spinners

Table 3 Payments for dependent spinners

$X+Y$	Probability
0	p_0
1	$p+q-2x$
2	x

The terms involving x cancel, in fact. Interestingly enough, for the variance an additivity relation holds only in case of independence of the games, i.e.

$$\text{var}(X+Y) = \text{var}(X) + \text{var}(Y) \text{ iff } P(X{=}1 \text{ and } Y{=}1) = P(X{=}1) \cdot P(Y{=}1),$$
$$\text{i.e. } x = p \cdot q.$$

***P*$_7$: Dependent Coins** Another example relates to drawing three coins (values 10p, 10p, 10p, 50p, 50p, 50p, 100p) from a bag with and without replacement (Borovcnik et al. 1991, p. 62). Calculations show that expected value is 3×40p, which is the same irrespective of replacement.

What is the Paradox? To calculate the probabilities, one has to know the result of the first draw in case the coins are not put back—the single draws turn to dependent random variables. Yet, for the calculation of the mean of the second draw one can neglect the result of the first. This is counter-intuitive. Furthermore, the design of the game without putting coins back prompts many people to reconstruct the situation personally. Speculating that the 100p coin can be drawn at first and decrease the net gain for the second is quite frequent resembling the sayings "the first wins" or "who dares to start wins".

Further Ideas This surprising result characterizes a substantial difference between probability and expectation. From the perspective of probability, the simplifying relation of linearity seems intuitively unacceptable if the random variables are dependent. However, for expectation there is symmetry; the probability of a specific coin being drawn at the first draw is the same as for the second or third try, even if coins are not replaced. Consistently, its individual contribution to each of the draws is the *same*. As this holds for all coins, the values are equal for all draws,

$$E(X_1) = E(X_2) = E(X_3) = 40\text{p},$$

whether coins are replaced or not. Thus the required expectation is 120p. Straightforward mathematical arguments support this reasoning. However, the historic situation was obscured by lengthy calculations, which outweighed its conceptual simplicity. The spinners and coins puzzles indicate that intuitions about expectation need to be discussed in teaching probability; these examples provide good stimuli.

In recent endeavours to enrich curricula and shift probability away from games of chance, risk is being taught in schools: situations are dealt with including the different outcomes *with* their related impact *and* their probabilities. Different options are then compared by their expected impact. For an individual the perception and evaluation of the probability of an outcome is influenced by its related impact. This interdependence is much increased if the probabilities are (very) small but the impact is enormous—as with screening programmes for preventing diseases like cancer. While risk extends the scope of teaching to an important field of application, it is inappropriate to introduce it right from the beginning. As the historical development has shown—the perception of probabilities, and the appreciation of a probability statement takes time to clarify; the notion of impact might obliterate the process of calibrating the feeling what a probability of e.g., 1/4 (or the proportion of 1 to 3) signifies for an event.

4 Relative Frequencies

The early attempts to explore probability were accompanied by the idea of frequency. The right way to count multiplicity was supported by the similarity of derived probabilities to the frequency of occurrence in repeated trials. Bernoulli (1713/1987) proved his law of large numbers and provided a justification to interpret probabilities as relative frequencies, which paved the way to many applications from life-tables to the behaviour of particles in physics.

Another move forward was Laplace's derivation of the central limit theorem (going back to preliminary results of de Moivre 1738/1967), which promoted the normal not only as a limiting distribution to the binomial but also as an element to formulate laws: laws in physics to describe the behaviour of entities at the microscopic level, but also laws to extract an estimation of unknown parameters from data. The focus of applications definitely turned towards empirical probabilities. The basis laid by Laplace was equi-probability and the principle of insufficient reason, which was criticized by empiricists like Venn. The time was ripe for probability as something like idealized relative frequencies (FQT).

The most serious attempt to classify relevant properties of relative frequencies in series of experiments axiomatically was made by von von Mises (1919), though it was dismissed as not sufficiently rigorous. The basic entities of his approach were *infinite series* of (theoretical) relative frequencies and some quite vague properties. One counter-argument to this approach was its complexity, while some contradictions were only repaired by Schnorr (1971). Kolmogorov's (1933) probability used the fundamental probability space and idealized relative frequencies for events instead of series of outcomes. It was universally acknowledged as a sound basis *and* also justified the interpretation of probability as relative frequency though there was an ongoing debate on repairing the direct approach by von Mises at the famous Geneva conference in 1937 (see the proceedings edited by Wavre 1938–1939).

Relative frequencies are based on independently repeating a random experiment, which is not always easy to define as shown by the exemplar paradoxes of the library problem and Bertrand's chord. It is startling to note that the conditions of randomness have to be operationalized when one would initially think that the experiment under scrutiny is random and its description is unambiguous. In both problems, the random selection of an object is operationalized in different ways thus confusing those who may think that random selection and the resulting relative frequencies must lead to a *unique* probability.

There are many idiosyncratic perceptions about how randomness manifests itself in repeated trials, which need to be addressed in the classroom. These perceptions underline the complexity of the concept of probability as the limit of relative frequencies. One irritating aspect refers to the patterns of a finite sequence of trials, which attract many to build up their own structure to continue the short-term behaviour of the frequencies—see Shaughnessy (2003), Konold (1989), or Borovcnik and Bentz (1991) for empirical teaching studies on such phenomena and strategies used. While the examples are puzzles in the sense of confusing problems, the features are also counter-intuitive with respect to personal thought and not due to a mathematical conception of probability as relative frequency.

Intuitively one might think that an experiment which is random has a unique formulation automatically, yet this is far from being true. This is made worse if the conception of probability is seen as a unique and almost physical property, like weight or length. So, if probability *is* nothing but the relative frequencies *in the long run*, how can it be that a problem gives rise to several experiments, which all depict the situation but lead to different relative frequencies and thus to different probabilities? If probability *were* only relative frequencies (of real objects), the situation would amount to a paradox, as illustrated by the library problem and the chord of Bertrand (1888).

P_8: Library Problem A book is selected randomly from the library. Determine the probability that it is written in English if there are 500 English out of 1,000 books in the library. A student has performed the experiments more than 2,000 times in going to the library randomly selecting a book; he reports a relative frequency close to 0.67. The librarian, on the other hand, has randomly chosen the book's index card from the search catalogue and got a relative frequency of 0.5. Why? (cf. Borovcnik et al. 1991, p. 60).

What is the Paradox? To choose randomly an element from a sample space seems to be unambiguous. Thus, following the steps the result should be the same for both. The paradox is that randomness has to be operationalized. There is no unique randomness, as the concept is bound to a *model* of the real situation. According to the model used, the relative frequencies differ and give different probabilities. The probabilities refer to the model rather than to the real situation while one may be surprised that there are various models to represent "random choice". One cannot "act" in the real situation ("behave" randomly) without using a model. This gives a clear hint that probability is a model entity rather than a physical property as it is related genuinely to a *model* of the world.

Assume that the library has a big and a small room, with a corridor in between. The student selects in two random steps: (i) throwing a coin to decide which room to enter; (ii) when in the room, selecting the book from the shelves from left to right according to a random number. Assume that there are $|E_1| = 410$ English books of a total of $|T_1| = 900$ books in the big room while for the small room the corresponding numbers are $|E_2| = 90$ and $|T_2| = 100$. A short calculation will show that, in fact, the student's random selection to get an English book by this selection is $61/90$ while using the card index yields 0.5.

If picking randomly is to be conceptualized by Laplace's approach then all the possible ways of selecting a book should yield the same probability. As various feasible (random) selection processes lead to different answers, the approach fails as long as one refers to the real situation instead of a model of it. The model is determined by the operational steps of selection.

***P*₉: Bertrand's Chord** An equilateral triangle is drawn in a circle with radius R and a line randomly drawn through the circle (Fig. 2). What is the probability that the segment s of the line in the circle is longer than the side a of the triangle? (Bertrand 1888, or Székely 1986, p. 43). Three possible solutions are given here (Figs. 3(a)–(c); cf. Borovcnik et al. 1991, p. 59).

(a) As the segment is uniquely determined by its mid-point M, we may focus on the position of M. If M is contained in the inner circle with radius R_1 with $R_1 = R/2$, we have $s > a$, otherwise $s \leq a$ (Fig. 3(a)). Hence

$$P(s > a) = \frac{R_1^2 \pi}{R^2 \pi} = \frac{1}{4}.$$

(b) We may compare the position of s on the diameter d perpendicular to s. If s falls within the interval I (see Fig. 3(b)), its length is greater than the length of a. As $|I| = R$, this yields

$$P(s > a) = \frac{|I|}{d} = \frac{1}{2}.$$

(c) As each segment s cuts the circle in P and Q, we may consider the angle β between s and the tangent t at Q in order to express the position of P in terms of β, which can lie in the range of (0, 180). If $60 < \beta < 120$, we have $s > a$ (Fig. 3(c)), thus

$$P(s > a) = \frac{60}{180} = \frac{1}{3}.$$

What is the Paradox? There should be a unique way to draw a chord in the plane. The puzzling issue is that the experiment is hard to perform in reality and the steps

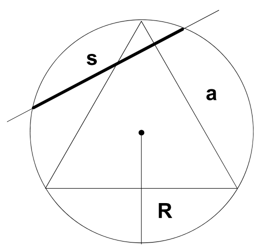


Fig. 2 Bertrand's chord: Line segment s and side a of the triangle to be compared

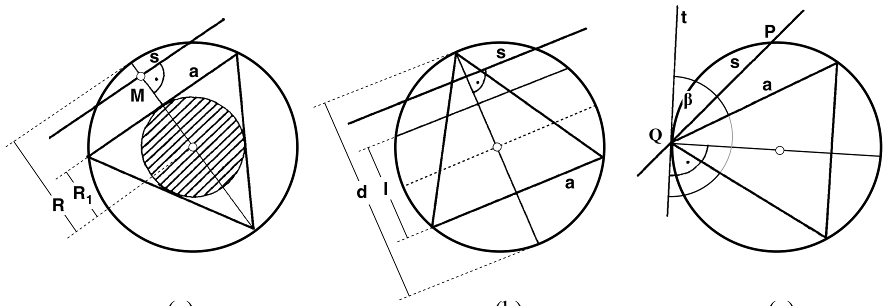


Fig. 3 (**a**)–(**c**): Bertrand's chord: Different possibilities to draw a random line

on how it is done have to be operationalized. In fact, the way the experiment is performed influences the result. Randomly drawing, in whatever way this may be defined, requires the use of Laplacean equi-probability on the possibilities which are open. Equi-probability refers only to the model and there is no guarantee that the right model is chosen. That implies that each of the three solutions represents chance and inherent equi-probability via its particular random generator. Is there more than one way of randomly determining a line? This reflects an intuitive conflict and yields a contradiction to the basic assumption of Laplace's definition; the word *randomly* is neither fully covered by this approach nor is it meaningful without reference to an actual generator of the events.

Further Ideas From today's perspective, there is no paradox. Probability is mathematically defined via the axioms and a stochastic experiment is described by different *models*, which may, of course, lead to different answers. The only question is which of the models in a real experiment delivers the better predictions. Only solution (b) fulfils the requirement of invariance to translations and rotations of the plane (see Székely 1986, p. 45 for a hint, or Palm 1983, for a full explanation), which is better in certain systems in statistical mechanics and gas physics. Nevertheless, the example shows that there is a problem worthy to discuss in the classroom situation. If not, then false intuitions may remain with children, and progress in learning probabilistic ideas and their application is hindered.

5 Personal Probabilities

There is little room (in APT and FQT philosophies) for the idea of probability as a judgement of a person about a statement or an event (SJT). This would make probability personal and subjective, rather than an objective concept. This view of probability is legitimatised by axioms on rational behaviour since de Finetti (1937) who states that probability does not exist, except as a personal idea: this approach is rejected by many as non-scientific. Where the focus is on the measurement of probabilities and the use of random experiments (which is an *idealization*), the critique against subjectivist probability is justified. The problem, however, is that with the exclusion of subjectivist probability other merits of this latter position are abandoned. The recurrent difficulties with conditional probabilities within a closed objectivist probability theory are a convincing argument that the idea of subjectivist probability is integral not only to people's intuitive reconstructions of mathematical concepts but also that a wider conception of probability is needed; this is especially true where events of low probability are concerned.

Despite the eminent role of independence, relevant intuitions are hard to clarify within mathematics. Some vague, nearly mystic arguments about "lack of causal influence" are used to back it up. It is interesting to note that subjectivists avoid these difficulties by replacing independence by *exchangeability*, an intuitively more accessible concept. The difficulties increase even with dependence, which is more than a simple complement to the notion of independence as there is a whole range of dependencies. Dependence is formalized by conditional probability, which is simply the "old" distribution restricted to the subspace determined by the conditioning event.

It should be no surprise that this mathematical approach gives rise to many difficulties in understanding. For example, for the dominant situation with equally likely cases, a reduction to a subspace cannot affect the equal probabilities—can it? On the contrary, for subjectivists, probability is a degree of belief and conditional probability is a basic notion which covers the intuitive idea of revising judgements as new information becomes available.

The Bayesian formula is a key tool and it is clear that for the final (posterior) probability judgement two ingredients have to be integrated, namely the prior probability of the states and the likelihood of the new information under the various states. The formula is so important that its clumsy appearance within Kolmogorov probability is changed into a more suitable and elegant mode for quicker re-evaluation of probabilities, which allows more direct insight on the relative importance of the influence factors (the prior probabilities and the likelihoods) and their impact on the final judgement. For this purpose, subjectivists often speak about probabilities in the form of odds or relative probabilities.

Carranza and Kuzniak (2009) analyse examples included in the curriculum with the conclusion that many deal with conditional probabilities and do not match the rest of the curriculum, which is oriented to the objectivist paradigm. According to objectivist paradigms, probability is strictly a property of objects which can be modelled differently mainly on the basis of available frequencies. The subjectivist paradigm relates probability to a judgement by a person who has some information available, which includes frequencies and qualitative information.

As the subjectivist position is criticized for being subjective (!), it was rejected as solution for a mathematical concept of probability. However, the Bayesian approach (SJT) is much closer to how many people think and can thus much better explain the part of conditional probabilities.

5.1 *Inverse Probabilities*

The following paradoxes show the difficulties in assimilating information in calculating probabilities. Intuitively many people do not believe that new information can change a probability. The application of Bayes theorem to calculate posterior probabilities is certainly complicated as shown by the furious international discussion of the Monty Hall problem (see Gigerenzer 2002). Here we present Bertrand's paradox and one relating to Father Smith.

***P*$_{10}$: Bertrand's Paradox** A cabinet has three boxes each with two drawers. Three gold and three silver coins are put into the drawers so that two boxes contain coins of the same kind and one the mixture. Randomly choose a box, then a drawer and open it; it is assumed that it contains a gold coin, which is denoted as event "G". Of interest is whether the box drawn first has coins of same type, which will be denoted as ST. After choosing the box but *before* the drawer is opened, there are 2 of 3 equally likely boxes, which yields $P(ST) = 2/3$ (Bertrand 1888; for an easy-to-play card version of the game, see Gardner 2006, p. 93; Gardner named it the "Three Card Swindle").

After seeing a gold coin, there remains 1 of 2 equally likely boxes, thus $P(ST|\text{“}G\text{”}) = 1/2$. *After* a silver coin is seen in the opened drawer, for symmetry, it holds that $P(ST|\text{“}S\text{”}) = 1/2$. If either a gold or a silver coin is in the opened drawer, the new probability is $1/2$. There is no need to look into the drawer, any result will decrease the probability to $1/2$, thus it *is* $1/2$. But probability of ST cannot be $2/3$ and $1/2$ at the same time (Figs. 4(a)–(b)).

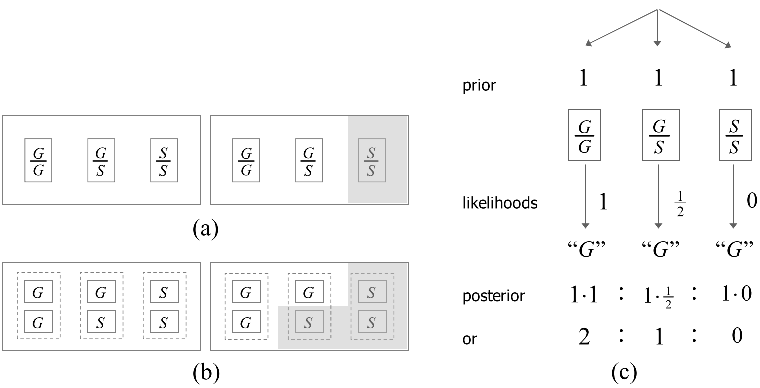


Fig. 4 Bertrand's paradox of drawers:
(**a**) Perceived random selection: 2 of 3 for *same type* before and 1 of 2 after seeing gold.
(**b**) Hidden random selection: 4 of 6 for *same type* before and 2 of 3 after seeing gold.
(**c**) Tree to combine priors and likelihoods

What is the Paradox? The focus on *equi-probable* objects (the boxes) leads to the trap. Logically, information "G" is used to reduce the possibilities—the $\boxed{\frac{S}{S}}$ box is eliminated. With equi-probability on the remaining boxes this yields $1/2$ for $\boxed{\frac{G}{G}}$ and $1/2$ for $\boxed{\frac{G}{S}}$; all in all this yields $1/2$ for same type and for mixed boxes. This argument leads to the paradox.

Seeing gold reduces the space, in fact, to the pure gold and the mixed box. However, the two-stage random selection leads to a hidden selection of the remaining coins as seen in Fig. 4(b), i.e. 2 out 3 gold coins lead to the pure gold box and thus for ST while the third leads to the mixed box. Thus the conditional probability remains at $2/3$ and the paradox is solved. This result can be confirmed by Bayes' formula. With equi-probable cases it is hard to think that this property of equal probabilities can be changed by new information. As, by definition, the conditional probability is simply a reduction of the present to a smaller space, how can this procedure change equal probabilities? The crux is that information "G" is used only on this logic base to reduce the space. Many people forget to discriminate between the other two boxes.

Further Ideas The mixed box has a conditional probability of $1/3$ confirming the hidden lottery argument. The situation can be generalized as it will enhance the *structure* of such situations.

The three boxes are perceived as hypotheses H_i and the evidence A as the result of the opened drawer. The hypotheses have a *prior* probability of $1/3$ each. The new or updated probabilities are calculated using Bayes' formula (the details are omitted):

$$P(H_i|A) = \frac{P(H_i \cap A)}{P(A)} = \frac{P(H_i) \cdot P(A|H_i)}{P(A)}.$$

Bayesians frequently use relative probabilities, so-called odds. For $P(E) = \frac{1}{6}$, the odds of E against its complement $\bar{E}$ are as $\frac{1}{6} : \frac{5}{6}$, or $1 : 5$. Odds are proportions but can freely be read as fractions. From odds of $1 : 5$, the probability is calculated back by $P(E) = \frac{1}{1+5} = \frac{1}{6}$, generally with odds of $a : b$, a probability of $P(E) = \frac{a}{a+b}$ is associated.

Comparing the updated probabilities of the hypotheses by odds yields (Fig. 4(c)):

$$\underbrace{\frac{P(H_i|A)}{P(H_j|A)}}_{\textit{posterior odds}} = \underbrace{\frac{P(H_i)}{P(H_j)}}_{\textit{prior odds}} \cdot \underbrace{\frac{P(A|H_i)}{P(A|H_j)}}_{\textit{likelihood ratio}}.$$

For Bertrand's cabinet this delivers a probability of $2/3$ from the posterior odds of $2 : 1$.

$$\frac{P(\boxed{\frac{G}{G}}|\text{"}G\text{"})}{P(\boxed{\frac{G}{S}}|\text{"}G\text{"})} = \underbrace{\frac{1}{1}}_{\textit{prior}} \cdot \underbrace{\frac{2}{1}}_{\textit{likelihoods}} = \frac{2}{1} = 2 : 1.$$

It is a deep-seated fallacy that the given information about a gold coin will leave the equi-distribution on boxes intact and that equal probabilities can be applied to the reduced sample space with the two remaining boxes—Bertrand (1888) favoured the wrong equi-probability of the two remaining boxes ending up with a paradox. There is also a reluctance to accept the results of Bayes' formula. Various didactical strategies to overcome this problem have been designed.

Freudenthal (1973) uses the technique of *implicit lotteries* (p. 590); the lottery on the boxes is symmetric, but the choice of the drawer is, by no means, symmetric. Falk and Bar-Hillel (1983) and Borovcnik (1987) suggest the *favour concept* which could intuitively clarify the higher estimate of the probability of the pure gold box as gold in the open division is circumstantial evidence for the box with two gold coins. Borovcnik and Peard (1996) suggest adapting mathematical formalism to fit better. The resulting view on Bayes' formula with odds connects objectivist and subjectivist conceptions. It gives a clearer view on the *structure* of the problem with prior possible states and evidence that leads to a new judgement of the probabilities. The value of an indication is represented by the relative likelihoods. The best is: to have evidence that has a high probability under one state and very small probabilities under the other states. Such an indication gives a clear new judgement. However, such situations are rare.

***P_{11}*: Father Smith and Son** Mr. Smith is known to have two children and various items of information may be analysed, leading to different posterior probabilities that he has two sons (this is the Two Children Problem of Gardner 1959, p. 51; cf. also Borovcnik et al. 1991, p. 64). The information is set out in Table 4 ((a) seen in town with a son, (b) visiting his home and randomly see a boy, (c) told eldest child is a boy, (d) told he has at least one boy, (e) told he prefers to go out with his son, (f) told he prefers to take his eldest child out, (g) told there are different probabilities for a boy or a girl to be at home).

What is the Paradox? It is confusing that information that seems to be similar or equivalent has a different impact on the probabilities. One has to judge how the information has been gathered before one can start to solve the problem.

Table 4 Different impact of the evidence on the posterior odds

Item	Information "*B*"	Posterior odds							Solution	Primitive result
		BB		*BG*		*GB*		*GG*	$P(BB\|$"B"$)$	
(a)	See in town	1	:	$\frac{1}{2}$	:	$\frac{1}{2}$	:	0	$\frac{1}{2}$	$\frac{1}{3}$
(b)	See at home	1	:	$\frac{1}{2}$	:	$\frac{1}{2}$	:	0	$\frac{1}{2}$	$\frac{1}{3}$
(c)	Eldest is boy	1	:	1	:	0	:	0	$\frac{1}{2}$	$\frac{1}{2}$
(d)	At least one boy	1	:	1	:	1	:	0	$\frac{1}{3}$	$\frac{1}{3}$
(e)	p prefers boys	1	:	p	:	p	:	0	$\frac{1}{1+2p}$	$\frac{1}{3}$
(f)	q prefers first	1	:	q	:	$1-q$	:	0	$\frac{1}{2}$	$\frac{1}{3}$
(g)	p_G, p_B at home								$\frac{2-p_B}{4-p_B-p_G}$	

Further Ideas In a structural view of the puzzle, the given information has to be linked to the possible states to re-evaluate them. The method of comparing prior and posterior probabilities of states can be applied to the information; but, as noted above, Bayes' theorem is complicated.

5.2 Conflicts with Logic

The theory of probability has a mathematical foundation, derived by logic. It is startling that reasoning with probabilities reveals a structure that appears to conflict with some of the rules of ordinary logic. This amounts to a puzzling situation as users erroneously expect all conclusions with probabilities to be in line with the following logical laws.

(a) *Transitivity* of logical reasoning. If A is bigger than B and B bigger then C, then one can conclude that A is bigger than C. Such a property signifies logical implication: If A implies B and B implies C, then—by transitivity—A implies C. This method establishes an important technique of mathematical proof.

(b) *Poof by exhaustion* or proof by cases. If a logical statement is true in either of two cases and these amount to all possibilities (and are disjoint), then the statement is true in all cases. That principle may be extended to 3 or more (countably infinite) cases, for example, if an equation $q(x) = 0$ holds for $x > 0$ (case 1), $x < 0$ (case 2) and $x = 0$ (case 3), it is true (for all real numbers x).

Probability statements conflict with these two principles, as shown by the following puzzles. As the logical relations seem quite natural, the clash is between properties of probability statements and intuitions. Nothing is wrong with probabilities thereby and nothing can be changed about these properties.

***P*₁₂: Intransitive Spinners** Suppose there are three spinners (Fig. 5). Which is the best to choose if two players compete and the higher number wins?

Player 1 chooses a spinner; player 2 chooses a spinner from the two remaining. There is no best choice for player 1; a short calculation shows that

$$P(S_1 > S_2) = 0.52, \qquad P(S_2 > S_3) = 0.61, \quad \text{and} \quad P(S_1 > S_3) = 0.25.$$

The second player can always find an alternative that is better. Player 1 is doomed to lose in this game. Recognizing this, it gets even more confusing that there *is* an

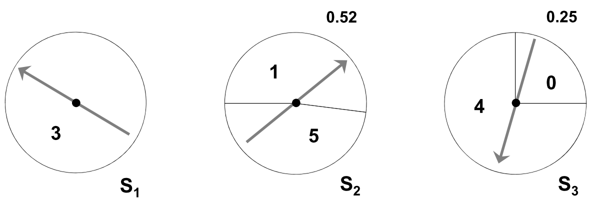


Fig. 5 Intransitive spinners

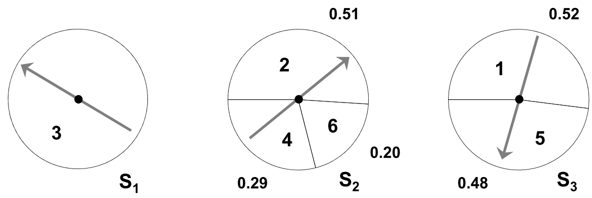


Fig. 6 Spinners from Blyth's paradox

order of the options with respect to losing (which is the complement to winning): To avoid losing too often, player 1 has a ranking on the spinners:

$$S_2 \succ S_3 \succ S_1$$

according to losing probabilities 0.52, 0.61 and 0.75 (against the best alternative).

What is the Paradox? If $P(S_i > S_j) > 0.5$ is interpreted as $S_i \succ S_j$ (and $\succ$ is used in the sense of "is better") then the properties of the spinners read as:

$$S_1 \succ S_2 \text{ and } S_2 \succ S_3 \text{ but } \textit{not}\text{—as transitivity would imply—} S_1 \succ S_3.$$

This is also puzzling in an everyday context: if A is better than B (in whatever respect) and B is better than C then, of course, A is better than C. A preference system that contradicts transitivity is counter-intuitive. Another puzzling feature is that with respect to winning there is no ranking for the spinners but to avoid losing too often there is a definite ranking. Is losing complementary to winning or not?

There is a high expectation that possibilities can be ranked according to some criterion and that ranking fulfils transitivity. The lack of transitivity in the choice illustrates that stochastics is different rather than a weak form of logic, as expressed by saying '... is true with probability p' instead of '... is definitely true'.

P_{13}: Blyth's Intransitive Spinners Blyth (1972) varies the situation (see Fig. 6): two or three players might enter the game. In fact, the optimal choice depends on the number of players. The calculations are left to the reader to explore because of restrictions of space, and similarly for the next puzzle.

P_{14}: Reinhardt's Single Spinner Reinhardt (1981) has a single spinner with several wheels with a similar puzzling result (Fig. 7).

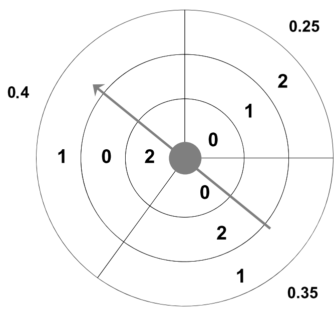


Fig. 7 Reinhardt's spinners

For a more detailed discussion of these spinners, see Borovcnik et al. (1991, p. 65). Puzzles like these can be motivating in the classroom situation. Usually they lead to a lively discussion of the applicability of such ideas; sporting situations such as in a football league are a suitable context. It certainly does happen that team A can beat team B but lose to team C, who are also beaten by team B. There are many intransitive situations in the real world that can cause confusion. A third player reversing the ranking of choice does occur.

P_{15}: Simpson's Paradox of Proportions In 1973, the following phenomenon was observed at the University of California at Berkeley: the overall admission rate of female applicants of 35 % was lower than that of their male colleagues, which amounted to 44 %. Females seemed to be less likely to gain admission. Searching for the reason for this sexual 'discrimination', it turned out that in some departments women actually had higher admission rates than men while most of the departments had similar rates for both (see Bickel et al. 1977). The following setting shows that admission rates can be higher for women than for men in *all* departments, and yet be lower for the whole university.

The situation is simplified by assuming there are only two departments. Green marbles represent admission, red marbles rejection, so, in Table 5, for example, 2 out of 5 females and 1 out of 3 males are accepted by department 1.

In both departments, the proportion of green marbles (admitted) is higher for females than for males, $2/5 > 1/3$ and $3/4 > 5/7$. But for the university as a whole, the reverse holds: $5/9$ for females admitted is lower than $6/10$ for males.

What is the Paradox? In department 1, the statement "women have a higher admission rate" is true. This holds also for department 2. Since, the two disjoint cases exhaust all the possibilities, why is the complementary statement true for the whole university? In what follows, a probabilistic framework for the situation will be established, first to show it deals with a *probabilistic* puzzle, second, to extend the analogy to logic.

Let F, M, G, R be the events female, male, green (admitted), red (rejected). Two urns are filled for the departments according to Table 5 and the experiment is drawing a ball within each department urn. Then the experiment is repeated with

Table 5 Illustrative data for Simpson's paradox

	Females		Males		All	
	Green	Red	Green	Red	Green	Red
Department 1	2	3	1	2	3	5
Department 2	3	1	5	2	8	3
University	5	4	6	4	11	8

Table 6 Some probabilities of interest within the departments and overall

	$P(F)$	$P(M)$	$P(R)$	$P(G)$	$P(G\|F)$
Department 1	0.625	0.375	0.625	0.375	0.400
Department 2	0.364	0.636	0.273	0.727	0.750
University	0.474	0.526	0.421	0.579	0.556

a university urn that is filled with all these balls. Some probabilities of interest are listed in Table 6.

In both departments, it holds that $P(G|F) > P(G)$, yet for the university the reverse relation holds as $P(G|F) < P(G)$. The reason for the intuitive clash lies in the *application rates* of males and females; females tend to apply to departments with low admission rates.

Further Ideas For logical implication $A \Rightarrow B$ the truth of statement A implies the truth of B. In analogy to this, a new (and weaker) relation between events is introduced. If one event increases the (conditional) probability of the other, i.e. if $P(B|A) > P(B)$, this is defined as $A \uparrow B$, in words, A *favours* B. Disfavouring, denoted as $A \downarrow B$ means that the conditional probability is smaller. With this concept, reasoning with probabilities is shown to differ from logical conclusions (Table 7).

Conditional probabilities are a subsidiary concept in the usual approach towards probability (either APT or FQT) and within the axiomatic approach there is no room for an investigation about the order or the direction of change of conditional probabilities. Such operations are, however, at the core of subjectivist theory (SJT) of probability, and it is important to integrate some elements from this position into teaching in order to enhance the underlying concepts, as advocated by Carranza and Kuzniak (2009) who supported their view by an analysis of teaching approaches.

The Simpson effect occurs in various contexts. Vancsó (2009) discusses an example with higher mortality in Mexico than in Sweden within all age classes (0–10, 10–20, etc.) but overall mortality being higher for Sweden than for Mexico. Underlying this version of the Simpson paradox is the fact that the Swedish population is much older while Mexico is a young country.

Table 7 Comparing the structure of logical reasoning and favouring

	Favouring	Logical implication
In case 1 it holds	$F \uparrow G$ and	$A \Rightarrow B$
In case 2 it holds	$F \uparrow G$	$A \Rightarrow B$
In all cases it holds	$F \downarrow G$ (in the example)	$A \Rightarrow B$ (generally)

6 Central Ideas of Probability Theory

In line with the intention of Kapadia and Borovcnik (1991), we present some central ideas in the theory of probability to provide a coherent treatment. Paradoxes indicate where a specific conception comes to an end and a reformulation of the terms is required to resolve the conflict. Puzzles show a divergence between official interpretations and private conceptions. Both paradoxes and puzzles though have limited implications. To develop a stable conception of notions, the mathematical *structure* is vital. What are the central ideas that link the concepts? We suggest the following set of ideas: independence and random samples; central theorems; standard situations; axiomatization. In order to reveal the value and role of axiomatization of probability, of course, a more detailed exposition of the underlying mathematics is important. To simplify here bears the risk of transmitting a limited picture of the theory and its potential for applications. The stronger mathematical demand in reading will pay off only afterwards by a deeper evaluation of the scope and limitations of probability.

6.1 Independence and Random Samples

The basic paradigm for probability is the experiment which can be repeated under essentially the same conditions and for which the outcome cannot be known beforehand with absolute certainty. This is modelled by a random variable X and the cumulative distribution F of X, i.e. $F(x) = P(X \leq x)$. The notion of independence extends naturally from events to random variables. For events A and B, independence means that

$$P(A \cap B) = P(A) \cdot P(B).$$

With the distribution function, the independence of random variables is defined by the condition:

$$F(x, y) = F_X(x) \cdot F_Y(y) = P(X \leq x) \cdot P(Y \leq y).$$

A sequence of independent random variables $X_1, X_2, \ldots$, each with the same distribution is called a *random sample* from that distribution and is denoted by

$$X_n \overset{iid}{\sim} X \sim F$$

(*iid* stands for independent and identically distributed). Such a series is a useful model for repeated observations, all independent and following the same distribution F.

The notion of a random sample is easily represented by a spinner which is independently spun several times. Another effective representation for sampling from a finite population is drawing balls from an urn, which is thoroughly mixed, prior to each draw.

6.2 Central Theorems

Historically, the development of the sum $H_n = X_1 + X_2 + \cdots + X_n$ has been investigated in special cases of the distribution of the single variables and H_n was binomially distributed. The deviations of H_n from their expected value tend to zero if n goes to infinity—Bernoulli's law of large numbers. Later, the probabilities of such deviations were asymptotically calculated, which led to the normal distribution and the central limit theorem. The assumption of independence was hidden in the binomial distribution.

Laws of large numbers show a convergence of the average of the random variables to a fixed value, while central limit theorems deal with the convergence of the standardized average from a sample to the normal distribution. Variations of the theorems cover different types of convergence of the average, other statistics like the median or the standard deviation derived from the series, specific distributions of single random variables, and restrictions leading to limiting distributions other than the normal distribution. We use modern notation to describe two theorems that became famous in the early mathematical development of the field, Bernoulli's law of large numbers and Laplace's central limit theorem.

Bernoulli's Law of Large Numbers Let A be an event of an experiment with $P(A) = p$, and X_i be a binary variable determined by an occurrence of A in independent repetitions, i.e.

$$X_i = \begin{cases} 1 & \text{if } A \text{ occurs at the } i\text{th trial,} \\ 0 & \text{if } A \text{ fails to occur at the } i\text{th trial,} \end{cases}$$

and let $H_n = X_1 + X_2 + \cdots + X_n$ be the absolute frequency of A in n trials, then for any positive real number ε

$$\lim_{n\to\infty} P\left(\left|\frac{H_n}{n} - p\right| \geq \varepsilon\right) = 0.$$

A generalization refers to the convergence of the mean of samples to the mean of the underlying population (cf. Meyer 1970, p. 246, or Çinlar 2011, p. 118).

If $X_1, X_2, \ldots, X_n$ are independent random variables from a common distribution with finite mean μ and variance σ^2 then, given $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| \geq \varepsilon\right) = 0.$$

The law of large numbers states that the mean of a sample will be close to the (unknown) mean of a distribution from which the sample was drawn, with a high probability provided that the sample size is sufficiently large and the selection process is random.

There is a strong version of both laws of large numbers stating that the set of infinite series which do not converge to the expected value of a single variable (i.e. p or μ) has a probability of zero (cf. Çinlar 2011, p. 122). This strong law of large numbers goes back to Borel (1909); its disputed status was not clarified until Kolmogorov's axiomatic foundation on the basis of measure theory.

Laplace's Central Limit Theorem The variable H_n (the absolute frequency) in Bernoulli's theorem will deviate from the expected value np in n independent trials so that it is a random variable with its own distribution. De Moivre considered a special case and Laplace found H_n to be approximately normal:

$$\lim_{n\to\infty} P\left(\frac{H_n - np}{\sqrt{np(1-p)}} \leq z\right) = \int_{-\infty}^{z} \varphi(t)\,dt$$

with $\varphi(t)$ being the probability density function of the standard normal distribution.

If the single random variables follow independently the same distribution as X (iid) and this distribution has a finite mean μ and a finite variance σ^2, the theorem would still hold, i.e.

$$\lim_{n\to\infty} P\left(\frac{H_n - n\mu}{\sqrt{n\sigma^2}} \leq z\right) = \int_{-\infty}^{z} \varphi(t)\,dt.$$

Thus, the central limit theorem is a natural basis to approximate the distribution of the mean from random samples in order to derive confidence intervals or statistical tests for the (unknown) mean. For a proof, see Meyer (1970, p. 250), or Çinlar (2011, p. 127).

Central Limit Theorem of Poisson Another limiting situation for the sum of variables, i.e. for $H_n = X_1 + X_2 + \cdots + X_n$ which is binomially distributed, was investigated by Poisson. Let the single summands be independent and binary with $P(X_i = 1) = p_n$ and consider a new parameter $\lambda = n \cdot p_n > 0$. For n tending to infinity and $X = \lim_n X_n$ it holds

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad \text{for } k = 0, 1, 2, \ldots,$$

i.e. the binomial distributions with this restriction converge to the Poisson distribution, which appears as the distribution of rare events as p_n tends to 0 as the product with n remains constant. For a proof, see Meyer (1970, p. 160), or Çinlar (2011, p. 137).

An example for modelling with the Poisson distribution is counting the atoms decaying in a specific period of time out of 1 kg of uranium U_{238}. The convergence above may be illustrated by the following ideal situation. If $n = 10$ raisins (points, events) should be distributed independently and randomly to 5 rolls (unit squares, observational unit), the number of raisins within a special roll is to be analysed. This corresponds to the random variable H_{10} with $X_i = 1$ if the ith raisin was attributed to it. Clearly, H_{10} is binomially distributed with parameters 10 and $1/5$ and expected number of raisins of $n \cdot p_n = 10 \cdot \frac{1}{5} = 2$. With $n = 100$ raisins and 50 rolls, the number of raisins in one special roll H_{100} is binomially distributed with 100 and $1/50$ and expected number of raisins of $n \cdot p_n = 100 \cdot \frac{1}{50} = 2$. In the latter situation, H_{100} is well approximated by the Poisson distribution with parameter $\lambda = 2$, which is the initial number of raisins per roll and is called the intensity of the process of distributing.

6.3 Standard Situations

Here we describe a few standard situations. We start with simpler processes such as those of Laplace and Bernoulli. We go on to more complex ideas of Markov chains and Brownian motion, typically studied at a university. It is remarkable that only a few distributions cover a wide range of applications. For modelling it is important to know the key idea behind the basic situation leading to that distribution. This explains the properties of the model and the modelled phenomenon.

Laplacean Experiments These are experiments where the equi-distribution is plausible on the basis of a physical symmetry. Conventional representations are spinners with equal sectors or urns filled with balls. For teaching, such experiments are useful to illustrate numerical probabilities to calibrate uncertainty.

Bernoulli Experiments The special case of a Laplacean experiment with two outcomes is known as a Bernoulli experiment. There are two different ways to explore this situation: to count "successes" in a specified number of trials leading to the binomial, or to wait for the next "success" leading to the geometric distribution.

Poisson Process Poisson experiments may be introduced as Bernoulli series in which the number of trials is high and the probability p is small. The Poisson process, however, describes a *genuine* random phenomenon of 'producing' events within time; heuristically, the process has to obey the following rules (see Meyer, p. 165).

(i) It does not matter when the observation of the process actually starts, the probabilities of various counts of events depend only on the *length* of observation.
(ii) For short periods the probability to have exactly one event is essentially the *intensity* λ of the process to produce events multiplied by the length of observation.
(iii) For short periods one may neglect the probability of two or more events.
(iv) Events occur independently in time.

The variable X which counts events in t units of time then follows a Poisson distribution:

$$P(X = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad \text{for } k = 0, 1, 2, \ldots;\ \lambda > 0.$$

Beside the probabilities of the number of events (Poisson), it is of interest to derive the probability for the waiting time for the next event (exponential distribution).

Elementary Errors Due to the central limit theorem, the distribution of an observed quantity X can be approximated by the normal distribution if it can be thought of as the result of a sum, i.e. $X = X_1 + X_2 + \cdots + X_n$. The analogy of small errors (the summands) contributing to generate the final quantity accounts for the ubiquity of the normal distribution. Historically, this reading of the central limit theorem has been a driving force and is still given in textbooks (Meyer 1970, p. 251).

Stochastic Processes The independence assumption is at the core of the random sample idea and also at the basis of central theorems like the law of large numbers and central limit theorem. There was another situation emerging from applications in physics that needed growing attention, where a probability measure was needed for an infinite dimensional Cartesian product: processes describing the change of a variable under scrutiny with progressing time. A slight shift in the description of the Poisson process will illustrate the conceptual change:

The variable X_t counts the number of events in the interval $(0, t]$. To be called a *Poisson* process, it has to fulfil the following conditions:

1. $X_0 = 0$. At the beginning the count starts at 0.
2. The process has stationary increments, i.e. the growth during time $(t, t+s]$ depends only on the length and not on the starting point of observation and its distribution depends only on s.
3. The process has independent increments, i.e. the growth in disjoint intervals is stochastically independent, i.e. for $t_0 < t_1 < t_2 < \cdots$ the increments $X_{t_1} - X_{t_0}$, $X_{t_2} - X_{t_1}, \ldots$ are independent.
4. With the exception of a set of zero measure, the trajectories $X_t(\omega)$ jump at most by 1 unit.

From the conditions one can conclude the following: The number of events in $(0, t]$ follows a Poisson distribution with parameter $\lambda \cdot t$, the waiting time for the next event is exponentially distributed with parameter λ, the location of any event in $(0, t]$ is uniformly distributed over this interval. The process is a Markov process with continuous time. Condition 2 corresponds to (i) and part of (ii) (the rest is not even required), 3 corresponds to (iv), while 4 corresponds to (iii).

An important Markov process was used as a tool in physics to describe the drifting of particles suspended in a fluid. A random walk in two dimensions means walking along the lattice of points with equal probabilities of 1/4 for continuing up, down, right, or left. By refining the grid more and more, and the central limit theorem, the following process—a so-called Wiener process—was motivated. Let X_t be the position of a particle at time t (usually this was a point in three-dimensional space).

1. $X_0 = 0$. The particle starts at the origin.
2. The process has stationary increments, i.e. the growth during time $(s, t]$ depends only on the length and not on the starting point of observation and its distribution depends only on the length $t - s$ and is, in fact, a normal distribution with expected value 0 and variance $t - s$.
3. The process has independent increments, i.e. the growth in disjoint intervals is stochastically independent, i.e. for $t_0 < t_1 < t_2 < \cdots$ the increments $X_{t_1} - X_{t_0}$, $X_{t_2} - X_{t_1}, \ldots$ are independent.
4. With the exception of a set of zero measure, the trajectories $X_t(\omega)$ are continuous functions of time.

For a similar formulation of the Poisson and Wiener process, see van Zanten (2010, p. 3). Most modern expositions like Çinlar (2011, p. 171), use the concept of martingales to define stochastic processes and are therefore less accessible. The phenomenon which was described in physics by such a model is Brownian motion. It is striking that at the time when such applications boosted the theory of thermodynamics, the foundations of probability was still not laid and probability was more or less justified by Laplacean equi-probability and interpreted as relative frequencies in long series of independent trials. But the reader should note that with the Markov processes above there was no independence in trials. And, of course, there was no firm foundation of a probability measure on an infinite-dimensional space as the trajectories of the process were the elements of the probability space. It was high time to solve the situation and in his 1900 address, Hilbert included the axiomatic basis of probability and mechanics as among the most urgent mathematical problems.

6.4 Kolmogorov's Axiomatic Foundation of Probability

Instead of the infinite random sequences of von Mises, Kolmogorov returns to a fundamental probability set which describes the potential of the experiment to produce outcomes in one trial. The question is how to define unambiguously the probability on the sample space and still have repeated independent experiments. The solution was to use measure theory to mathematize probability like other measures such as area or weight, and add the concept of independence between different trials subsequently. Repeated trials are modelled by the sample space which is built from the Cartesian product of the sample space of single experiments.

The Axioms Kolmogorov (1933) developed a system of axioms for the special case of a finite sample space $S = \{x_1, x_2, \dots, x_m\}$. Instead of using the power set of S (which is feasible here), he refers to a field $\mathcal{F}$ of events (an algebra of events) as the domain for a probability function. A field $\mathcal{F}$ thereby is a system of subsets of S, which has the property that the set operations of union, intersection and complementation (finitely) applied on elements of $\mathcal{F}$ always yield a set that belongs to $\mathcal{F}$ (i.e. the field is closed under the usual set operations). His axioms are (p. 2):

(I) $\mathcal{F}$ is a field of sets.
(II) $\mathcal{F}$ contains the set S.
(III) To each set A in $\mathcal{F}$ a non-negative real number $P(A)$, the probability of A, is assigned.
(IV) $P(S)$ equals 1.
(V) If A and B have no element in common, then $P(A \cup B) = P(A) + P(B)$.

The conditional probability is defined for a fixed event A with $P(A) > 0$ as

$$P_A(B) := \frac{P(A \cap B)}{P(A)}$$

with the justification that the function P_A fulfils the axioms. Two events A and B are defined as independent if

$$P(A \cap B) = P(A) \cdot P(B),$$

which is shown as equivalent to the following relations if both events have a positive probability:

$$P_A(B) = P(B) \quad \text{and} \quad P_B(A) = P(A).$$

At this point, Kolmogorov proceeds (p. 14) to define probability measures on *infinite* spaces by adding one more axiom, the so-called continuity of a probability measure:

(VI) For a strictly decreasing sequence of events $A_1 \supset A_2 \supset \cdots \supset A_n \supset \cdots$ of $\mathcal{F}$ with $\bigcap_n A_n = \emptyset$ the following equation holds $\lim_n P(A_n) = 0$ for $n \to \infty$.

Ensuring this axiom, the countable additivity is proved as a theorem, i.e. it holds that

$$P\left(\bigcup_n A_n\right) = \sum_{n=1}^{\infty} P(A_n) \quad \text{for any sequence } A_n \text{ of } \mathcal{F} \text{ with } A_i \cap A_j = \emptyset \text{ for } i \neq j;$$

i.e., the additivity holds for any *sequence* of pairwise disjoint events.

In the rest of his seminal work, Kolmogorov refers to the system $\mathcal{F}$ as a σ-field (σ-algebra), i.e. he requires the field to be closed also under *countably* infinite applications of the usual set operations by the following argument:

> Only in the case of [σ] fields of probability do we obtain full freedom of action, without danger of the occurrence of events having no probability. (p. 16)

Modern representations of Kolmogorov's axioms prefer to relocate his first two axioms into the denotation of a probability measure P as a real function on a σ-algebra $\mathcal{F}$, i.e.

$$P : \mathcal{F} \to \mathbf{R}$$

and refer to the σ-additivity instead of the continuity so that the axioms read as the following conditions on the function P:

(A_1) $P(A) \geq 0$ for any event A from $\mathcal{F}$.

(A_2) $P(S) = 1$ for the whole sample space S.

(A_3) If $A_0, A_1, \ldots$ is a sequence of mutually exclusive events from $\mathcal{F}$, then $P(\bigcup_{n=0}^{\infty} A_n) = \sum_{n=0}^{\infty} P(A_n)$.

The first two conditions mean that probabilities are non-negative and that certainty is characterized by a value of 1. The substantial condition of *σ-additivity* embodied in A_3 means that, mathematically, probability is a measure. It may be regarded in some respect as analogous to 'area', 'mass', or 'weight', measures which also share the additivity property.

Kolmogorov (p. 1) believes that

> The theory of probability, as a mathematical discipline, can and should be developed from axioms, in exactly the same way as Geometry and Algebra [...] all further exposition must be based exclusively on these axioms, independent of the usual concrete meaning of these elements and their relations.

The special choice of a set of axioms has deep consequences on semantics. The axioms are the foundation of the theory and are simultaneously at the interface between theory and reality. The basic axioms could be considered as models of intuitive ideas of probability to be sharpened by the theory. This might be thought of as the historical genesis of ideas in general and also in the sense of how ideas settle down in an individual's learning.

Distribution Functions In his paper, Kolmogorov uses the concept of a (cumulative) distribution function extensively (p. 19). This only makes sense if the concept fully characterizes a probability measure. Therefore, before turning to examples of probability functions on the space of real numbers, Kolmogorov (p. 16) proves an abstract extension theorem, which states that a probability measure on a field $\mathcal{F}$ can be uniquely extended to the smallest σ-field $[\mathcal{F}]$, which contains $\mathcal{F}$. The importance of that theorem cannot be overestimated as in the real numbers the domain of a probability function, the so-called Borel σ-algebra $\mathcal{B}$ of events is precluded from any intuitive access. But, luckily, it is possible to focus on a simple generating system of it which is built up of specific intervals $(a, b]$ (a and b could be real numbers or $\pm\infty$) and their finite unions, which form a field $\mathcal{F}$. Even more, a simpler generating system suffices as there are more general extension theorems that do not require the structure of a field on the generator.

$$\mathcal{C}^* = \{(a, b] \mid a, b \in \mathbf{R}\} \quad \text{or} \quad \mathcal{C} = \{(-\infty, x] \mid x \in \mathbf{R}\}.$$

That means, of any probability measure P on the Borel σ-algebra $\mathcal{B}$ on $\mathbf{R}$ it suffices to know the values of P on sets of $\mathcal{C}$. Or, conversely, the pre-probabilities fulfilling the axioms on sets of $\mathcal{C}$ uniquely determine a probability measure P on the Borel sets $\mathcal{B}$.

The complicated story with the Borel sets and σ-algebras has its origin in the following theorem. There can be no probability measure P fulfilling all the axioms for *all* subsets of $\mathbf{R}$. A contradiction can be derived if it is assumed that a probability can be attributed to *all* subsets with all the named properties. To avoid this, the domain of the function P has to be restricted to a true subset of the power set of $\mathbf{R}$. The natural structure of all *admissible* sets (for probability) is that of a σ-algebra, which is a system of subsets that is closed under the usual set operations, countably often applied in any order. The Borel sets $\mathcal{B}$ serve this purpose perfectly.

Probability Measures on Infinite-Dimensional Spaces What distinguishes the theory based on this set of axioms from measure theory is the concept of independence, which is part of a fundamental definition, but not, interestingly, part of the axioms. This independence relation is the key assumption of fundamental theorems like Bernoulli's law of large numbers. Such theorems established a link from the structural approach to the frequentist interpretation and thus contributed to the immediate acceptance of Kolmogorov's axioms within the scientific community.

More complications arise in the case of infinite sample spaces for single trials as the sets of the form $E_1 \times E_2$ (which are known as *cylinder sets*) are only a small part of all subsets of $S_1 \times S_2$. In practice, events in the combined experiment may not be of this special form, e.g. in spinning a pointer twice, consider the event 'position of the trials differ by more than $\pi/4$'. This complication is not a conceptual difficulty of probability as a phenomenon to be modelled but is linked to specific aspects of mathematics. For applications it is fortunate that assigning probabilities to cylinder sets is sufficient to uniquely determine an extension of this assignment to probabilities of all events. For infinite-dimensional spaces, the trick was used to start from cylinder sets defined on a finite number of coordinates (cf. Kolmogorov, p. 27).

Lebesgue Integral There is another unifying element in Kolmogorov's fundamental paper, namely the use of integrals for calculating probabilities and expected values. The distribution function F_X uniquely defines a probability measure P_X on the Borel sets. Probabilities and expected values may be written as integrals as follows

$$P_X(A) = \int_A dF_X(x) \quad \text{and} \quad E(X) = \int x\, dF_X(x),$$

which are Lebesgue–Stieltjes integrals. This unified the theory of probability of discrete and continuous distributions. For practical needs these integrals can be evaluated as ordinary sums or—in case of intervals as events, i.e. $A = (a, b)$, as integrals; for the bulk of applications the Lebesgue integral is not even required in evaluating the integrals involved, as the Riemann integral suffices for the most important distributions:

$$P_X(A) = \begin{cases} \sum_{i \in A} p_i & X \text{ with discrete probabilities } p_i = P(X = i), \\ \int_a^b f_X(x)\, dx & X \text{ with absolutely continuous density } f_X = \frac{d}{dx} F_X, \end{cases}$$

$$E(X) = \begin{cases} \sum_i i p_i, \\ \int x f_X(x)\, dx. \end{cases}$$

These ideas are well beyond school level. Yet, it is important to be aware of and remember this theory. It forms the deep foundation on which probability has been built.

It is important to note that the many deep results like the central limit theorem were derived before a sound axiomatic basis had been established and they retained their validity and importance after the axiomatization. What was different is the prestige probability gained as a scientific discipline, which then attracted many young researchers to the field. Furthermore, axiomatization paved the way to probability distributions on infinite-dimensional spaces and the field of stochastic processes, which revolutionized not only physics but many other fields like financial mathematics. For example, the price of an option in the financial market is derived by the solution of a stochastic differential equation of a stochastic process similar to the one described above.

7 Conclusions

Modern expositions of probability such as Çinlar (2011) have reached an elegance of mathematical standard, which is sometimes in sharp contrast to the philosophical framework. The situation resembles somehow the early culmination of development with Laplace working on the central limit theorem but expressing a naïve determinism that probability is only for those ignorant of the causes. The general philosophical debate such as von von Plato (1994) in the context of physics shows dramatically that probability without a firm philosophical footing misrepresents its scope in limits as well as in reach.

The standard paradigm is to interpret probability initially as equal possibilities and then as the limit of relative frequencies—or as relative frequencies from samples large enough. All the concepts of inferential statistics from the objectivist position heavily draw on this paradigm. This way was paved by Kolmogorov's own views on his axioms justifying the frequentist conception of probability. The reaction from the subjectivist position was fierce, led by de Finetti who ironically noted, in capital letters that "PROBABILITY DOES NOT EXIST" (1974, p. x); he sees probability as a way to *think about* the world. But their mathematical exposition—grounded on axioms of rational behaviour and rational updating of probabilistic information via Bayes' formula—normally uses a different terminology and a unified exposition of probability respecting both subjectivist and objectivist ideas has not been published, yet.

Urgent topical problems on the objectivist side are that small probabilities are growing in importance, yet data is missing or is highly unreliable, or originates from qualitative knowledge. Probability is often used more in the sense of a scenario which means that probabilistic models are used as a heuristic to explore reality instead of finding the best-fitting model and determining the "best" solution relative to it for the real problem under scrutiny (see Borovcnik and Kapadia 2011). Another source of confusion is an adequate understanding of statistical methods that reminds one of the historical problems to understand the puzzling examples or paradoxes on the sole basis of an objectivist probability.

What remains of probability if it is deprived of its main interpretation as relative frequencies is hard to tell. Some key properties such as the additivity of expected values or the key idea behind any distribution, which sets the scene for a structure of situations—a structure that goes beyond and behind the fact that the relative frequencies do fit and can be modelled by it, are illustrated in Borovcnik (2011). The link to relative frequencies remains too dominant. The conception of a degree of belief and how to revise it by new data gives guidance to understand the notions and related methods much more easily. The discussion of the paradoxes shows that such ideas enhance understanding.

A well-balanced exposition covering subjectivist and objectivist probability seems unlikely. Barnett (1982) marked a promising step into this direction with a comparative analysis of the positions but this remains an isolated project within the statistical community despite the fact that the theory based on Kolmogorov's axioms could serve as a common language. Barnett (1982, p. 69) notes that these axioms are a "mathematical milestone" in laying firm foundations; they remain, however, a "philosophical irrelevance" in terms of explaining what probability really is.

Returning to paradoxes and fallacies, they can be entertaining. They raise class interest and motivation. Discussion of these ideas can help to

– analyse obscure or complex probabilistic situations properly;
– understand the basic concepts better;
– interpret formulations and results more effectively;
– balance and shift between different interpretations of probability;
– educate probabilistic intuition and reasoning more firmly.

The misconceptions in the examples show that probabilistic intuitions seem to be one of the poorest among our natural and developed senses. Perhaps, this is a reflection of the desire for deterministic explanation. People have great difficulty in grasping the origins and effects of chance and randomness: they search for pattern and order even amongst chaos. Or, it may be due to an education too restricted to one perception of probability. The examples above illustrate the gap between intuition and mathematical theory, particularly because stochastic reasoning has no empirical control to revise inadequate strategies. Paradoxes and puzzles highlight these difficulties as signs of a cognitive conflict between an intuitive level of reasoning and formalized, mathematical arguments. In a paradox, the 'objective' side is inadequate though intuitively straightforward, whereas in a puzzle the objective side is adequate but intuitively inaccessible. Empirical research in using paradoxes in teaching is limited, though a promising start has been made by Vanscó (2009) with trainee teachers; this now needs to be replicated in schools. Our own long and varied experience leads us to assert that planned discussion of paradoxes and puzzles fosters individual conceptual progress of children and students in learning probability.

References

Barnett, V. (1982). *Comparative statistical inference* (1st ed. 1973). New York: Wiley.

Bayes, T. (1763). An essay towards solving a problem in the Doctrine of Chances. *Philosophical Transactions of the Royal Society*, *53*, 370–418. Reprinted in E. S. Pearson, & M. G. Kendall (1970), *Studies in the history of statistics and probability* (Vol. 1, pp. 131–154). London: Griffin.

Bernoulli, J. (1713/1987). *Ars conjectandi*. Basel: Impensis Thurnisiorum, Fratrun. Translation of 4th part by N. Meunier. Rouen: Institut de Recherche sur l'Enseignement Mathematique. German translation by R. Haussner (1899), *Wahrscheinlichkeitsrechnung*. Leipzig: Engelmann.

Bernoulli, D. (1738/1954). Specimen theoriae novae de mensura sortis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, *5*, 175–192; English translation: L. Sommer, Exposition of a new theory on the measurement of risk. *Econometrica*, *22*, 23–36.

Bertrand, J. (1888). *Calcul des probabilités*. Paris: Gauthier-Villars.

Bickel, P. J., Hammel, E. A., & O'Conell, W. J. (1977). Sex bias in graduate admissions: data from Berkeley. In W. B. Fairley, & F. M. Mosteller (Eds.), *Statistics and public policy*. Reading: Addison-Wesley.

Blyth, C. R. (1972). Some probability paradoxes in choice from among random alternatives. *Journal of the American Statistical Association*, *67*, 366–381. With comments by Lindley, D. V., Good, I. J., Winkler, R. L., and Pratt, J. W.

Borel, É. (1909). Les probabilités dénombrables et leurs applications arithmétique. *Rendiconti del Circolo Matematico di Palermo*, *27*(1), 247–271.

Borovcnik, M. (1987). Revising probabilities according to new information—a fundamental stochastic intuition. In R. Davidson, & J. Swift (Eds.), *Proceedings of the second international conference on teaching statistics* (pp. 298–302). Victoria: University of Victoria. Online: www.stat.auckland.ac.nz/~iase/publications/icots2/Borovcnik.pdf.

Borovcnik, M., & Bentz, H.-J. (1991). Empirical research in understanding probability. In R. Kapadia, & M. Borovcnik (Eds.), *Mathematics education library: Vol. 12*. *Chance encounters: probability in education* (pp. 73–105). Dordrecht: Kluwer Academic.

Borovcnik, M., & Kapadia, R. (2011). Modelling in probability and statistics—key ideas and innovative examples. In J. Maaß, & J. O'Donoghue (Eds.), *Real-world problems for secondary school students—case studies* (pp. 1–44). Rotterdam: Sense Publishers.

Borovcnik, M., & Peard, R. (1996). Probability. In A. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education* (pp. 239–288). Dordrecht: Kluwer Academic.

Borovcnik, M., Bentz, H.-J., & Kapadia, R. (1991). A probabilistic perspective. In R. Kapadia, & M. Borovcnik (Eds.), *Mathematics education library: Vol. 12*. *Chance encounters: probability in education.* (pp. 27–71). Dordrecht: Kluwer Academic.

Carranza, P., & Kuzniak, A. (2009). Duality of probability and statistics teaching in French education. In C. Batanero et al. (Eds.), *Joint ICMI/IASE study: teaching statistics in school mathematics* (5 pp.). Online: www.stat.auckland.ac.nz/~iase/publications/rt08/T1P2_Carranza.pdf.

Çinlar, E. (2011). *Probability and stochastics*. Berlin: Springer.

d'Alembert, J. (1754). Croie ou pile. In D. Diderot, & J. d'Alembert (Eds.), *Encyclopédie ou Dictionnaire raisonné des sciences, des arts et des métiers* (Vol. 4, pp. 512–513). Paris: Société des Gens de Lettres.

David, F. N. (1962). *Games, gods and gambling*. London: Griffin.

de Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, *7*, 1–68. Foresight: Its logical laws, its subjective sources (Chaps. 1–4 translated by H.E. Kyburg Jr.). In S. Kotz, & N.L. Johnson (1992), *Breakthroughs in statistics. Vol. I. Foundations and basic theory* (pp. 134–174). New York: Springer.

de Finetti, B. (1974). *Theory of probability*. New York: Wiley. Translated by A. Machi, & A. Smith.

de Moivre, A. (1738/1967). *The doctrine of chances* (2nd ed., fuller, clearer, and more correct than the first). London: Woodfall. Reprint of 3rd ed. 1967. New York: Chelsea. 1st ed. 1718. London: Pearson.

Falk, R., & Bar-Hillel, M. (1983). Probabilistic dependence between events. *Two-Year College Mathematics Journal*, *14*, 240–247.

Fermat, P., & Pascal, B. (1679). Correspondence 1654. In P. Fermat (Ed.), *Varia Opera Mathematica*. Toulouse: Joannem Pech. English translation in F. N. David (1962), *Games, gods and gambling* (pp. 229–253). London: Griffin.

Freudenthal, H. (1973). *Mathematics as an educational task*. Dordrecht: Reidel.

Galilei, G. (ca. 1613–1623, 1962). *Sopra le Scoperte dei Dadi*. Fragment on dice. Published under the title "Considerazione sopra il Giuoco dei Dadi" in A. Favaro (1890–1897) (Ed.), *Le Opere di Galileo Galilei* (Vol. VIII, pp. 591–594). Firenze: Tipografia La Barbera. English translation in David, F. N. (1962), *Games, gods and gambling, Appendix 2* (pp. 192–195, translated by E. H. Thorne).

Gardner, M. (1959). *The Scientific American book of mathematical puzzles & diversions*. New York: Simon & Schuster.

Gardner, M. (2006). *Aha! A two volume collection: aha! Gotcha aha! Insight*. Washington: Math. Assoc. of America.

Gigerenzer, G. (2002). *Calculated risks: how to know when numbers deceive you*. New York: Simon & Schuster.

Huygens, C. (1657). De ratiociniis in ludo aleae. In F. v. Schooten: *Exercitationes matematicae*, Leyden: Elsevirii.

Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, *XLVII*, 263–291.

Kapadia, R., & Borovcnik, M. (Eds.) (1991). *Mathematics education library: Vol. 12. Chance encounters: probability in education*. Dordrecht: Kluwer Academic.

Kolmogorov, A. N. (1933). *Ergebnisse der Mathematik, 2. Band, Heft 3. Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin: Springer. Reprinted 1977 (62 pp.). English translation 1956: *Foundations of the theory of probability*. New York: Chelsea.

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

Laplace, P. S. (1812/1951). Essai philosophique sur les probabilités. *Journal de l'École Polytechnique*, *VII/VIII*, 140–172. English translation: *A philosophical essay on probabilities* (extended version). New York: Dover.

Maistrov, L. E. (1974). *Probability theory. A historical sketch*. New York: Academic Press.

Meyer, P. L. (1970). *Introductory probability and statistical applications* (2nd ed.). Reading: Addison-Wesley.

Pacioli, L. (1494). *Summa de arithmetica, geometria, proportioni et proportionalitá*. Venedig: Paganino de' Paganini.

Palm, G. (1983). Wo kommen die Wahrscheinlichkeiten eigentlich her? *Der Mathematik-Unterricht*, *29*(1), 50–61.

Price, R. (1764–1765). A demonstration of the second rule in Bayes' essay. *Philosophical Transactions of the Royal Society of London*, *54*, 296–297; *55*, 310–325.

Reinhardt, H. E. (1981). Some statistical paradoxes. In A. P. Shulte, & J. R. Smart (Eds.), *Teaching statistics and probability* (pp. 100–108). Reston: National Council of Teachers of Mathematics.

Schnorr, C. P. (1971). *Lecture notes in mathematics: Vol. 218. Zufälligkeit und Wahrscheinlichkeit. Eine algorithmische Begründung der Wahrscheinlichkeitstheorie*. Berlin: Springer.

Shaughnessy, J. M. (2003). Research on students' understanding of probability. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 216–226). Reston: National Council of Teachers of Mathematics.

Székely, G. J. (1986). *Paradoxes in probability and mathematical statistics*. Dordrecht: Reidel.

Vanscó, Ö. (2009). Parallel discussion of classical and Bayesian ways as an introduction to statistical inference. *International Electronic Journal in Mathematics Education*, *4*(3), 291–322. Online: www.iejme.com/032009/main.htm.

Venn, J. (1888). *The logic of chance* (3rd ed.). London: Macmillan. 1st ed., published 1866.

von Mises, R. (1919). Grundlagen der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, *5*, 52–99.

von Plato, J. (1994). *Creating modern probability: its mathematics, physics and philosophy in historical perspective*. Cambridge: Cambridge University Press.

Wavre, R. (Ed.) (1938–1939). Colloque consacré à la théorie des probabilités, Proceedings of the Geneva conference 1937; published in 8 Vols. Paris: Hermann. *Actualités Scientifiques et Industrielles*, *734–740*, 766.

Zanten, H. V. (2010). *An introduction to stochastic processes in continuous time*. Leiden: University of Leiden. Online: www.math.leidenuniv.nl/~spieksma/colleges/sp-master/sp-hvz1.pdf.

Three Approaches for Modelling Situations with Randomness

Andreas Eichler and Markus Vogel

Abstract Three different approaches to the concept of probability dominate the teaching of stochastics: the classical, the frequentistic and the subjectivistic approach. Compared with each other they provide considerably different possibilities to interpret situations with randomness. With regard to teaching probability, it is useful to clarify interrelations and differences between these three approaches. Thus, students' probabilistic reasoning in specific random situations could be characterized, classified and finally, understood in more detail. In this chapter, we propose examples that potentially illustrate both, interrelations and differences of the three approaches to probability mentioned above. Thereby, we strictly focus on an educational perspective.

At first, we briefly outline a proposal for relevant teachers' content knowledge concerning the construct of probability. In this short overview, we focus on three approaches to probability, namely the classical, the frequentistic and the subjectivistic approach. Afterwards, we briefly discuss existing research concerning teachers' knowledge and beliefs about probability approaches. Further, we outline our normative focus on teachers' potential pedagogical content knowledge concerning the construct of probability. For this, we discuss the construct of probability within a modelling perspective, with regard to a theoretical perspective on the one side and with regard to classroom activities on the other side. We further emphasize considerations about situations which are potentially meaningful with regard to different approaches to probability. Finally, we focus on technological pedagogical content knowledge. Within the perspective of teaching probability, this kind of knowledge is about the question of how technology and, especially simulation, supports students understanding of probabilities.

A. Eichler (✉)
University of Education Freiburg, Kunzenweg 19, 79117 Freiburg, Germany
e-mail: andreas.eichler@ph-freiburg.de

M. Vogel
University of Education Heidelberg, Keplerstraße 87, 69120 Heidelberg, Germany

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_4,

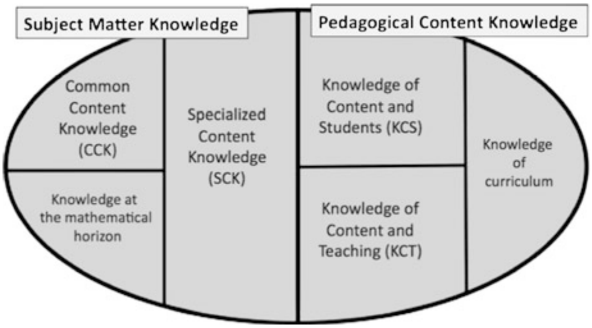


Fig. 1 Different kinds of knowledge which teachers should have according to Ball et al. (2008)

1 Introduction

> The success of any probability curriculum for developing students' probabilistic reasoning depends greatly on teachers' understanding of probability as well as much deeper understanding of issues such as students' misconceptions [...]. (Stohl 2005, p. 345)

With these words Stohl (2005) begins her review on teachers' understanding of probability. The first part of this quote seems to be self-evident. However, probability is known as a difficult mathematical concept yielding "counterintuitive results [...] even at very elementary levels" (Batanero and Sanchez 2005, p. 241). Even to grasp the construct of probability itself lasted centuries from elaborating the classical approach (Pascal 1654, cited in Schneider 1988), the frequentistic approach (von Mises 1952) to an axiomatic definition by Kolmogoroff (1933). Moreover, the subjectivistic approach that fits Kolmogoroff's axioms but in some sense negate the existence of objective probabilities (De Fineti 1974; Wickmann 1990; Batanero et al. 2005a), complement the set of very different approaches to the construct of probability. Particularly when the different approaches of probability that dominate the stochastics curricula (Jones et al. 2007) are analysed a crucial question arises (Stohl 2005): What must teachers know about the construct of probability?

In this chapter, we will deduce an answer to this question from a pure educational perspective. For this we use a currently widely accepted model of Ball et al. (2008). Following this model, teacher should have knowledge of the content (CK: content knowledge; Fig. 1) and as part of CK, a knowledge that is "not typically needed for purposes other than teaching" (SCK: specialized content knowledge; Ball et al. 2008, p. 400). Teachers further should have pedagogical knowledge which combines knowledge about students (KCS), curriculum, and content and teaching (KCT) which "combines knowing about teaching and knowing about mathematics" (Ball et al. 2008, p. 401).

Some of the relevant domains of stochastical knowledge (cf. Ball et al. 2008) are discussed in this volume, e.g. the horizon content knowledge represented by philosophical or mathematical considerations. In addition, the knowledge we gained about students (KCS) is reviewed extensively elsewhere (e.g. Jones et al. 2007). For this reason we focus on aspects of the teachers' content knowledge (KC) with a specific perspective on the teachers' specialized content knowledge (KCS). We center on the teachers' pedagogical knowledge and, in particular, on the teachers' knowledge of content and teaching (KCT) from a normative perspective. This normative perspective we expand partly referring to results of empirical research concerning students' knowledge (Jones et al. 2007) and teachers' knowledge (Stohl 2005).

2 Three Different Approaches to Probability (Content Knowledge)

We restrict the discussion about different approaches of probability to a brief overview of knowledge that teachers should have since deeper analyses exists already (e.g. Borovcnik 1992 or Batanero et al. 2005a). Highlighting the central ideas of each approach from their underlying philosophical points of view (including the historical context of their emerging) conduces to set out the crucial aspects, which teachers have to consider when they think about relevant conditions of teaching probability. Thus, there is a base for argumentation when outlining pedagogical ideas of teaching probability by means of specific random situations.

2.1 The Classical Approach

Although there were many other well-known historic figures dealing and thinking about probability, the classical approach is namely connected with Pierre Simon Laplace and his groundbreaking work "Essai Philosophique sur les Probabilités" from 1814. One very essential point of this work is the definition of probability, which we today know about as "Laplace-probability": "Probability is thus simply a fraction whose numerator is the number of favorable cases and whose denominator is the number of all cases possible." (Laplace 1814/1995, p. ix cited by Batanero et al. 2005a, p. 22) The underlying theoretical precondition of this definition is the equiprobability given for all possible outcomes of the random process being regarded. This modelling principle of having no reason for doubting on equiprobability can mostly be maintained by using artificial, symmetrical random generators like a fair die, an urn containing balls, which are not distinguishable from each other with regard to be drawn by chance, or a symmetrical spinning wheel and treating them as being theoretically perfect. Because of these preconditions the term "a priori" is connected to the classical approach to probability measurement (e.g. Chernoff 2008; Chaput et al. 2011). "A priori" means that, together with the underlying theoretical hypothesis of equiprobability being aligned before, there is no need to conduct the experiment with a random generator. In fact, probabilities that refer to a random generator can be calculated deductively without throwing a die or spinning a wheel. From a mathematical point of view, Laplace-probability is only applicable in a finite set of possible outcomes of a random process; in case of an infinite set it is not possible to adopt this approach of probability.

2.2 The Frequentistic Approach

With its strong preconditions the classical approach of probability measurement is an approach rooted in a theoretical world allowing for predictions within the empirical world. What is its advantage on the one side is its disadvantage on the other side: many random phenomena of the empirical world (more exactly said, phenomena explained as being caused by random) are not suitable to be purely theoretically regarded as a set of equiprobable outcomes of a random generator. For example, in case of throwing a pushpin: Before throwing the pushpin the first time, no theoretical arguments can be reasonably deduced so that the probability for landing on the pushpin's needle can exactly be predicted. But already during the time of Laplace the empirical fact was well known that the frequencies of a random experiment's outcome are increasingly stabilizing the more identical repetitions of the random experiment are taken into account. This empirical fact, called "empirical law of large numbers", served as a basis to estimate an unknown probability as the number around which the relative frequency of the considered event fluctuates (Renyi 1992). Since the stabilization is observable independently by different subjects, the frequentist approach was postulated as being an objective approach to probability. It was Bernoulli who justified the frequentist approach of probability by formulating a corresponding theorem: Given that there is a random experiment being repeated enough times, the probability that the distance between the observed frequency of one event and its probability is smaller than a given value can approach 1 as closely as desired. Von Mises (1952) tried to take a further step forward and defined probability as the hypothetical number towards which the relative frequency tends during the stabilization process when disorderly sequences are regarded. This definition was not without criticism: It was argued that the kind and conditions of the underlying limit generation was not sufficient with regard to a theoretical point of view of the mathematical method.

From a practical point of view, it was criticized that this empirical oriented approach is only useful for events which allow for repeating the underlying experiment as often as needed, at least in theory. However, within phenomena of the real world this is often neither possible nor practicable. Apart from that, in principle, no number can be fixed to ensure an optimal estimation for the interesting probability. Nevertheless, the frequentistic approach enriched the discussion about probability because of its empirical rootedness. Because of this the term "a posteriori" is connoted with the frequentistic approach to probability (e.g. Chernoff 2008). According to the frequentistic approach, it is necessary to gain data (or rather relative frequencies) concerning the outcomes of a random generator for estimating corresponding probabilities. Within this experimental approach, probability can be interpreted as being a physical magnitude which can be measured, at least principally, as exactly as needed.

2.3 The Subjectivistic Approach

Within the subjectivist approach, probability is described as a degree of belief, based on personal judgement and information about a situation that includes objectively no randomness. One example to illustrate this fact concerns HIV infection (cf. Eichler and Vogel 2009). For a specific person the objective probability to be infected with HIV is 0 or 1. Thus, the infection of a specific person comprises no randomness. However, a specific person could hold an individual degree of belief about his (or her) status of infection (may be oriented to the base rate of infection). This degree of belief can change by information, e.g. a result of a diagnostic test. Thus, subjectivistic probabilities depend on several factors, e.g. the knowledge of the subject, the conditions of this person's observation, the kind of event whose uncertainty is reflected on, and available data about the random phenomena.

The subjectivist point of view on probability is closely connected with the Bayesian formula being published in the late eighteenth century. The Bayesian formula allowed for revising an "a-priori"-estimation of probability by processing new information and for estimating a new "a-posteriori" probability. In general, the subjectivistic approach calls the objective character postulated by the frequentist approach or the classical approach into question. With this regard Batanero et al. (2005a, p. 24) state: "Therefore, we cannot say that probability exists in reality without confusing this reality with the theoretical model chosen to describe it." Thus, following the subjectivistic approach, it is not possible to treat probability as being a physical magnitude and to measure it accordingly in an objective way. Beyond this fundamental difficulty, Wickmann (1990) finds fault with the frequentistic approach in being very limited with regard to its applicability within real world problems because the theoretical precondition of an unlimited number ($n \in \mathbb{N}$) of an experiment's repetitions is not given within limitedness of the real world. Mostly, in reality this number is comparably small being not sufficiently large for an objective estimation of probability.

3 Teachers' Knowledge About Probability

In the last section, we have not differentiated between common content knowledge (CCK) and a specific teacher's specialized content knowledge (SCK). Nevertheless, we interpret knowledge concerning the distinction and connections between "a priori" probability (classical approach) and "a posteriori" probability (frequentist approach) as well as the combination of "a priori" and "a posteriori" probability (subjectivist approach) as knowledge teachers must have. In particular, the knowledge of connection between the mentioned three approaches of probability has been often claimed (e.g. Steinbring 1991; Riemer 1991; Stohl 2005; Jones et al. 2007).

However, research has yielded only scarce results of what teachers actually know about probability and about how to teach probability in the classroom (Shaughnessy 1992; Jones et al. 2007). Begg and Edwards (1999) reported weak mathematical knowledge of 34 practicing and pre-service elementary school teachers concerning the construct of probability (classical and frequentistic approach). Further, Jones et al. (2007) reported (citing the research of Carnell 1997) on difficulties which 13 secondary school teachers had when they were asked to interpret conditional probabilities that are a prerequisite for teaching the subjectivistic approach of probability. Stohl (2005) and Jones et al. (2007) do not mention further research that underlines the difficulties teachers seem to have when dealing with probabilities.

Similarly, research in stochastics education has provided scarce results about how teachers teach the three mentioned approaches of probability and, thus, about the teachers pedagogical content knowledge. A quantitative study in Germany referring to 107 upper secondary teachers showed that all of them teach the classical approach (Eichler 2008b). While 72 % of these teachers indicated to teach the frequentist approach 27 % indicated to teach the subjectivist approach. This research yielded isolated evidence that although the European classroom practice showed a strong emphasis to probability (Broers 2006) the challenge to connect all three approaches of probability is still not achieved (Chaput et al. 2011). Further, a study of Eichler (2008a, 2011) showed that even when teachers include the frequentistic approach in their classroom practice the way to teach this content can considerably differ. While one teacher emphasizes the frequentist approach by experiments with artificial random generators (urns, dice, cards) another teacher uses real data sets. Doing so, the latter teacher conforms to the paradigm of teaching the frequentist approach mentioned by Burril and Biehler (2011). Finally, Stohl (2005) reported that teachers' knowledge about connecting different approaches of probability may be due to a limited content knowledge. However, Dugdale (2001) could show that preservice teachers benefit from computer simulation to realize connections between the classical and the frequentist approach of probability.

Although the empirical knowledge about teachers' content knowledge and pedagogical content knowledge concerning the three approaches of probability is low, in the following we use existing research to structure a normative proposal for teaching these three approaches of probability. Of course, such a proposal needs development along a coherent stochastics curriculum across several years including the development of the underlying mathematical competencies. Referring to this proposal, we firstly outline three situations of throwing dice (artificial random generator) that provide the three approaches of probability. We connect these three situations using a modelling perspective. Afterwards, we focus on situations which do not base on artificial random generators to emphasize the modelling perspective on the three approaches of probability. Finally, we discuss computer simulations that potentially could facilitate students' understanding of the three approaches of probability.

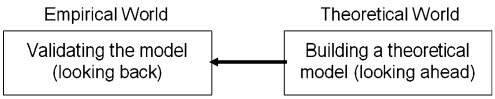


Fig. 2 Modelling structure with regard to the classical approach

4 Probability Measurement Within the Modelling Perspective

An important competency for the thinking and teaching of probability is the competency of modelling (e.g. Wild and Pfannkuch 1999). Instead of differentiating distinct phases of modelling (e.g. Blum 2002), we restrict our self to the core of mathematical modelling, i.e. the transfer between the empirical world (of data) and the theoretical world (of probabilities). The discussion about the different types of probability measurement shows that each approach of probability is touching both the theoretical world and the empirical world as well as the interrelationship of these worlds.

One very essential characteristic of the concept of probability is its orientation towards future—probabilities are used to quantify the possibility of occurring a certain event's outcomes which still hasn't happened. The situational circumstances and the available information (including the knowledge of the probability estimating person) are decisively determining how the process of quantifying probability is possible to go on. In principle, three different scenarios of mutual processes of looking ahead and looking back can be distinguished when the modelling process is considered as transfer between the empirical and the theoretical world (cf. Vogel and Eichler 2011). We will illustrate each kind of modelling by means of possible classroom activities. In doing so, we use rolling dice to validate these modelling processes quantitatively. Concerning these classroom activities, we restrict the discussion on secondary schools, although in primary schools in some sense the teaching of probability is also possible.

4.1 Modelling Structure with Regard to the Classical Approach

With regard to the classical approach, the modelling structure principally includes the following four steps (Fig. 2):

1. *Problem:* Defining and structuring the problem
2. *Looking ahead:* Building a theoretical model based on available information and making a prediction on base of this theoretical model (*looking ahead*; theoretical world)
3. *Looking back:* Validating this prediction by producing data (for example, by doing simulations) based on this theoretical model and analysing these data in a step of reviewing (*looking back*)
4. *Findings:* Drawing conclusions and formulating findings

A classroom activity concerning the mentioned four steps is as follows:

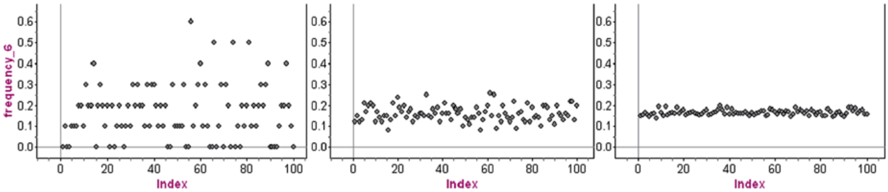
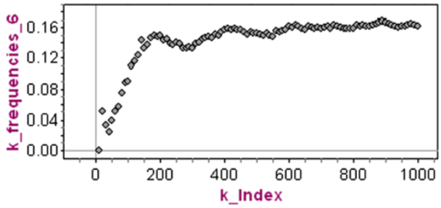


Fig. 3 Relative frequency of the six referring to series of throwing the die 10, 100 and 1000 times

Fig. 4 Cumulated frequencies of the six in series of ten throws



Problem: The students are given ordinary dice. They have to predict the relative frequency of the six referring to the future 10, 100 and 1000 throws. Afterwards the students have to validate their predictions.

Looking ahead: Most of the students, who reached secondary school, presumably, do not doubt that the chance given for each side is "a priori" $1/6$ (for example, by analysing the symmetrical architecture of the cube). Given that the students have experienced the variability of statistical data (Wild and Pfannkuch 1999), they might predict the frequency to get a six by $1/6$ accepting qualitatively small differences between the real frequency and the predicted frequency (which corresponds to the "a priori" estimated probability) of $1/6$.

Looking back: To be able to decide if the predictions are appropriate, the students have to throw the die. This can be conducted by hand or by computer simulation, following to the law of large numbers: "the more the merrier." The empirical data represent at most qualitatively the $\frac{1}{\sqrt{n}}$-law (Fig. 3).

Findings: The intervals of frequencies of the six get smaller when the series of throws gets bigger. Thus, the frequencies of the six will fit the model of $1/6$ better the more throws are performed. In another similar experiment, these findings lead to the empirical law of large numbers when the relative frequencies of series of throws are cumulated (Fig. 4).

All the findings mentioned above are based on the theoretical model of the probability of the six "a priori." Thus, coping with the given problem starts by considerations into the theoretical world of probabilities that have to be validated into the empirical world of data.

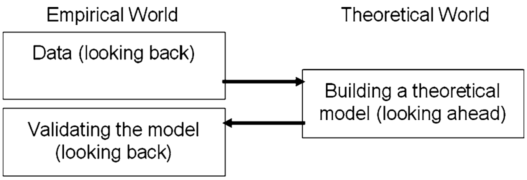


Fig. 5 Modelling structure with regard to the frequentistic approach

4.2 Modelling Structure with Regard to the Frequentistic Approach

With regard to problems following the frequentistic approach, the modelling structure includes the following five steps (Fig. 5):

1. *Problem:* Defining and structuring the problem
2. *Looking back:* Analysing available empirical data of the interesting phenomena for detecting patterns of variability that might be characteristic
3. *Looking ahead:* Building a theoretical model based on available information and extrapolating these patterns for purposes of prediction or generalization of this phenomena
4. *Looking back:* Validating this prediction by producing new data (for example, by doing simulations) based on this theoretical model and analysing these data in a further step of reviewing
5. *Findings:* Drawing conclusions and formulating findings

A classroom activity concerning the mentioned five steps is as follows:

Problem: The students get a cuboid die (in Germany, a so-called "Riemer-Quader" Riemer 1991) and have to predict the relative frequency of the numbers referring to the future 10, 100 and 1000 throws. Afterwards the students have to validate their predictions.

Looking back: Although it is possible to hypothesize about a model concerning the cuboid die, i.e. the probability distribution, at the end nothing more can be done except throwing the cuboid die. Before that some students tend to estimate probabilities according to the relative ratio between the area of one side of the cuboid and the whole surface (there might be also other reasonable possibilities). However, such a model will not endure the validation. Thus, students get a model by throwing the cuboid die as many times as possible to get reliable information about the relative frequency of the different outcomes using the empirical law of large numbers. The students might be confident with this approach: When they follow the same procedure using the ordinary die (see above), they can get an empirical affirmation because they know the probabilities' distribution of the ordinary die.

Looking ahead: The empirical realizations are the base for estimating probabilities for each of the cuboid's sides. With regard to the modelling process it is crucial to remark that the relative frequencies cannot be simply adopted as probabilities. In the given problem, the geometrical architecture of the cuboid die gives a reason for assigning the same probabilities to each pair of opposite sides of the cuboid. For example, if the distribution of the relative frequencies of the cuboid is given by $h(1) = 0.05$; $h(2) = 0.09$; $h(3) = 0.34$; $h(4) = 0.36$; $h(5) = 0.11$; $h(6) = 0.05$, then the students should think about what to do with these unsymmetrical empirical results. One appropriate consideration obviously seems to be to take the mean of two corresponding sides and to estimate the probabilities based on the two corresponding frequencies (but not adopting one of the frequencies). Accordingly, with regard to the example mentioned before, a reasonable estimation for the distribution of probabilities could be: $P(1) := P(6) := 0.05$; $P(2) := P(5) := 0.10$; $P(3) := P(4) := 0.35$. In this way, the students deduce patterns of probabilities from the available data and readjust them with regard to theoretical considerations within the modelling process.

Looking back: The validation process in the case of the cuboid die equals to the validation process of the ordinary die.

Findings: Since long series of throwing the ordinary die potentially yield in confidence concerning the stabilization of relative frequencies, relative frequencies of throwing the cuboid die might serve as an appropriate basis to estimate the probability distribution. An estimation of probabilities is not done by simply adopting the relative frequency. This is an essential finding that a partly symmetrical random generator like the cuboid die provokes. Thus, the cuboid die seems to be more appropriate than a random generator like the pushpin for which the estimation of a probability is potentially equal to the relative frequency.

4.3 Modelling Structure with Regard to the Subjectivistic Approach

With regard to problems following the subjectivist approach, the modelling structure is not restricted. Thus this structure could potentially include user-defined steps since the number of alterations between the theoretical world and the empirical world is not defined (Fig. 6).

1. *Problem:* Defining and structuring the problem
2. *Looking back:* Subjective estimation (theoretically or empirically based, not in general exactly distinguishable) based on given information
3. *Looking ahead:* Coming to a provisional decision and making a subjective but theoretically based prediction

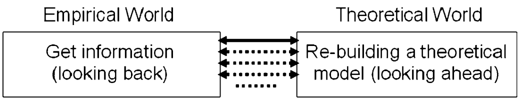


Fig. 6 Modelling structure with regard to the subjectivistic approach

4. *Looking back:* Inspecting the first and provisional decision by generating and processing new empirical data

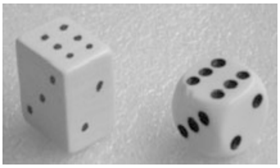
   ...

   (Now, the process of modelling alternates between the steps 3 and 4 as long as an appropriate basis to decide between two or more hypothesis is reached. The assessment of being appropriate is subjective, the prediction is the more objective the more information is processed.)

   ...

n. *Findings:* Drawing conclusions and formulating findings

A classroom activity concerning the mentioned steps is as follows:



Problem (cf. Riemer 1991; Eichler and Vogel 2009): The students are playing a game along the following rules: There are small groups of students each with one referee. The referee has two dice, an ordinary die (O) representing an equal distribution of probability and a cuboid die (R—"Riemer-Quader") with the distribution: 1–0.05; 2–0.10; 3–0.35; 4–0.35; 5–0.10; 6–0.05 (derivation see above, Sect. 4.2). The dice are covered for the other members of the referee's group. The referee has covertly chosen one of the dice and throws it repeatedly. The other students try to find out which die the referee has chosen by processing information about the diced spots and by estimating step-by-step the probabilities for the die covertly been thrown.

Looking ahead: Because there are two dice and because the students have no information about the referee's preference, the first estimation ("a priori") could yield $P(R) = 0.5$ for the cuboid die and $P(O) = 0.5$ for the ordinary die.

Looking back: The referee informs the students about having thrown a 1.

Looking ahead: Qualitatively, this outcome gives evidence for the ordinary die since this die should produce a 1 more often than the cuboid die ($P(1|O) > P(1|R)$). By using the Bayesian formula, this information can help to get a new estimation of probability for the referee's die-choice at the beginning:

$$P(O|1) = \frac{P(1|O) \cdot P(O)}{P(1|O) \cdot P(O) + P(1|R) \cdot P(R)} = \frac{\frac{1}{6} \cdot \frac{1}{2}}{\frac{1}{6} \cdot \frac{1}{2} + \frac{1}{20} \cdot \frac{1}{2}} = \frac{10}{13} \approx 0.769.$$

Correspondingly there is a new estimation a posteriori with $P(O) := 0.769$ and $P(R) := 0.231$.

Results of throwing the unknown die:
1—2—5—6—6—3—6—1
$P(O) \approx 0.998$; $P(R) \approx 0.002$

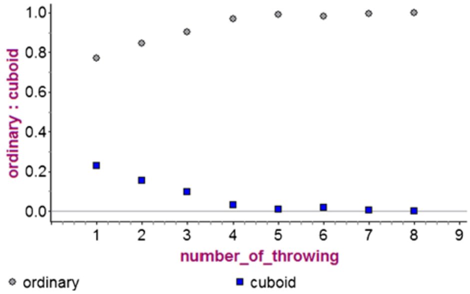


Fig. 7 Development of estimating probabilities which results from successively applying Bayesian formula

Looking back: The referee informs the students about having thrown a 2. The a posteriori-estimation of the first round of the game simultaneously is the a priori-estimation of the following second round.

Looking ahead: Qualitatively, this outcome gives again evidence for the ordinary die since this die should produce a 2 more often than the cuboid die ($P(2|O) > P(2|R)$). Using the Bayes formula again results in:

$$P(O|2) = \frac{P(2|O) \cdot P(O)}{P(2|O) \cdot P(O) + P(2|R) \cdot P(R)} \approx 0.847.$$

Correspondingly, there is a new estimation a posteriori with $P(O) = 0.847$ and $P(R) = 0.153$.

Looking alternating back and ahead: A series of further throws is shown in Fig. 7.

This Fig. 7 clearly shows that the step-by-step processing yields at the end a good basis for the decision.

Findings: The processing of information according to the Bayesian formula could yield an empirical based decision concerning different alternative hypotheses. The decision is not necessarily correct. Given the results above and given that the referee used covertly a third die (a cuboid die for which the 1 and 6 are on the biggest sides, and the 3 and 4 on the smallest sides), the decision for the ordinary die would actually be false in an objective sense. But within the subjective perspective, it is the optimal decision based on available information.

4.4 Some Principles of Modelling Concerning Probability

With regard to the three situations mentioned above, it is to remark that, in general, there is no fixed rule how to model a problem, it depends on kind and circumstances of the problem, on the one side, and on the problem solver's abilities and knowledge, on the other side (cf. Stachowiak 1973). The basic idea of modelling is to desist from the complexity of the original question and to concentrate on aspects which are considered as being useful for getting an answer. This is at the core of modelling: A model is useful for someone within a specific situation to remedy specific demands (cf. Schupp 1988). Thus, all models cannot be judged by categories of being right or wrong, they have to be judged by their utility for solving a specific problem (cf. Box and Draper 1987).

However, concerning modelling there is a fundamental difference among using a model of objective probabilities (classical and frequentistic approach) and subjectivistic probabilities. Since the former approach could validate by collecting new data, in the latter approach the validation is integrated in every step of readjusting a probability for a decision to one of different (and potentially incorrect) models.

In the cases when data can be interpreted as being empirical realizations of random processes, the probability modelling process could be considered within the paradigm of Borovcnik's (2005) structural equation for modelling data (cf. Eichler and Vogel 2009): an unknown probability is assumed to be immanent in a random process (cf. Sill 1993) as well as to be appropriate to quantify a certain outcome of this process. Regardless of the different approaches mentioned above, the modelling process in general aims to get a reliable estimation for this unknown probability. The estimation is a *modelling pattern* approximating the unknown probability. It means: This estimation results from pattern which the modelling person detects by analysing the random process and its different outcomes. According to Stachowiak (1973), there is always a difference between a model and the modelled original. In the light of Borovcnik's (2005) structural equation, this difference and the modelling pattern mentioned above sum up to the unknown probability. Thus, it could be highlighted that not the unknown probability itself is captured by the modeller but what he (or she) perceives, what he already knows about different types of random processes and accordingly reads into this process.

5 Problem Situations Concerning Three Approaches of Probability

Following Greer and Mukhopadhyay (2005), for many years a well-known prejudice against teaching stochastics is the fact that stochastics is applied only by formula-based calculating within the world of simple random generators. This world is represented by random processes like throwing dice as discussed above as well as spinning wheels of fortune, throwing coins, etc. However, the world of gambling is merely connecting the theoretical world of probability with the empirical world if real data are regarded representing the statistical part of teaching stochastics. With reference to Steinbring and Strässer (1981) Schupp (1982) states: "Statistics without probability is blind [...], probability without statistics is empty" (translated by the authors). Burril and Biehler (2011, p. 61) claimed that "probability should not be taught data-free, but with a view towards its role in statistics."

At a first view, our "dice world" considerations of Sect. 4 seem to contradict these statements. Nevertheless, our concern of emphasizing the three approaches to probability using throwing dice is to be interpreted in a slightly different perspective. Within a pedagogical point of view, the "dice world" must not be avoided at all: This world also contains data (e.g. empirical realizations of random processes like throwing dice) and provides situations which are reduced from complex aspects of real world contexts. Thus, such situations might facilitate students' understanding of different approaches to probability. However, remaining in such a "dice world" must be avoided: Although this artificial world fulfils a crucial role for the students' mathematical comprehension of probability, teaching probability must be enriched by situations outside the pure world of gambling. Thus, the students can deepen their understanding of probability by reflecting on its role in applying statistics. By this, probability can become more meaningful to the students. So, changing the citation of Schupp (1982) a little, it can be said that probability without accounting for realistic situations and real data stays empty for students.

Our statement is underlined from psychology with regard to theories of situated learning and anchored instruction. "Situations might be said to co-produce knowledge through activity. Learning and cognition, it is now possible to argue, are fundamentally situated." (Brown et al. 1981, p. 32) Essential characteristics (among others) of this situational oriented approach are interesting and adequate complex problems to be mastered as well as realistic (not necessary real) and authentic learning surroundings containing multiple possible perspectives on the problem (cf. Mandl et al. 1997). By following the goal to prevent the so-called *inert knowledge*, the research group *Cognition and Technology Group at Vanderbilt* (1990) developed and applied in practical oriented research the theory of *anchored instruction*. The central characteristic of this approach is its narrative anchor by which the interest of the learners should be produced and the problem should be embedded in the context so that situated learning can take place. Within this perspective, we suggest that emphasizing only the "dice world" would not meet the demands of a perspective of situated learning or anchored instruction briefly outlined above.

Considering the different situations that could be regarded concerning approaches of probability measurement and different roles data are playing during the modelling process, we differentiate three kinds of problem situations (Eichler and Vogel 2009):

- Problem situations which contain all necessary information and represent a stochastic concept like an approach to probability in a strongly reduced context (that is usually not given within the complex circumstances of real world problems) we call *virtual problem situations*.
- Problem situations which demand analysing a situation's context that is more "authentic" and provide a "narrative anchor" (see above), but still provide demands that do not aim to reconstruct real world problems we call *virtual real world problem situations*.
- Problem situations that include the aim to reproduce real problems of a society we call *real world problem situations*.

In the following, we exemplify three situations of the latter two kinds of problems representing the three approaches of probability outside the world of gambling.

5.1 A Virtual Problem Situation Concerning the Classical Approach

A computer applet concerning the so-called Cliffhanger Problem could be seen as an example (source: http://mste.illinois.edu/activity/cliff/ (12.2.2012)).

A long day hiking through the Grand Canyon has discombobulated this tourist. Unsure of which way he is randomly stumbling, $1/3$ of his steps are towards the edge of the cliff, while $2/3$ of his steps are towards safety. From where he stands, one step forward will send him tumbling down. What is the probability that he can escape unharmed?

With regard to the underlying mathematical structure, this problem situation is one variation of the so-called random walk problems. Defined by the question for the unknown probability, it is comparable to the task to find out what the probabilities of a fair dice, or rather a fair coin, are.

A normative solution for this situation represents the classical approach of probability. Given the positions measured in steps from the edge by 0, 1, etc., and denoting probabilities for falling down at different positions by $p_1, p_2, \ldots$, the probability of falling down starting at the position 1 is represented by: $p_1 = 1/3 + 2/3 p_2$. Further, the probability of moving from position 2 to position 1 is the same as moving from position 1 to position 0 (i.e. falling down), namely p_1. Thus, the probability p_2 of falling down at the position 2 equals to p_1^2, i.e. $p_2 = p_1^2$. Combing these two equations leads to the quadratic equation $p_1 = 1/3 + 2/3 p_1^2$, which is solved by $p_1 = 0.5$ (the second solution 1 apparently does not make sense within the given situation).

Those students who are able to find a theoretical solution afterwards have the opportunity to check their considerations by using a simulation which is enclosed in the computer application. Thus, (virtual) empirical data could help validate the theoretical considerations.

However, the situation must not be analysed by using the classical approach of probability. Students who do not get along well with the classical approach have the opportunity first to use the computer application for simulating and thus, for getting a concrete image of the problem situation and its functioning as well as an idea what the solution could be. In this way, their attempt to find a theoretical explanation is guided by their empirically funded assumption of about 0.5 being the right solution.

As already mentioned above, there is no fixed rule how to solve such problem situations: starting with theoretical considerations followed by data based validating or starting with data based assumptions followed by theoretical explanations or even multiple changing between theoretical considerations and data based experiences, it is a modelling question by aiming to make decisions and predictions within the given computer based problem situation.

The Cliffhanger Problem is structurally equivalent to the problem of throwing an ordinary die. It provides also artificial simplifications within a given virtual sit-

Averages

blue	brown	green	orange	red	yellow
3.1	3.1	2.9	3.1	3.0	2.8

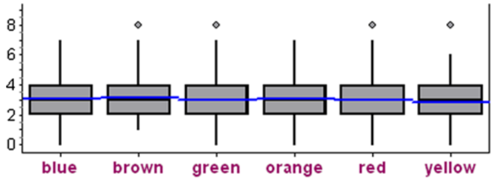


Fig. 8 Statistics concerning 100 opened candy bags

uation, but emphasizes approaches to probability in a situation beyond the "dice world" mentioned above. Students have the opportunity to apply their stochastic knowledge and mathematical techniques according to a situated learning arrangement prepared by the teacher. Thus, students do not only learn something about the situation's stochastic content but also about techniques of problem solving. In this sense, such virtual problem situations could allow for speaking of learning about stochastic modelling in a *mathematically protected area under laboratory conditions*.

5.2 A Virtual Real World Problem Situation Concerning the Frequentistic Approach

Explore the distribution of colors in the candy bags.

The candy bags contain six different colors. By focusing on only one color, the complexity could be reduced to a situation which in a teacher's forethought could allow for modelling by using a binomial distribution (assuming that the mathematical pre-conditions are given). Provided that there is no further information available (e.g. about production process of the candy bags), only a theory-based approach to the problem seems not to be given. The empirical character of the problem situation indicates that data is needed (Wild and Pfannkuch 1999) before reliable hypothesizing about the colors' distribution is possible.

Opening of 100 candy bags has produced data which are represented in statistics of Fig. 8.

To found a theoretical model on these descriptive analyses it is necessary to name the underlying assumptions: For example, with regard to color each candy bag is modelled as being filled independently from the others. Further, the probability of drawing one candy with a specific color is assumed not to change. Beyond that a further modelling assumption is that all candy bags contain the same number of candies. Accepting these modelling assumptions and thus, assuming there is an equal distribution of all six colors founding an underlying theoretical model, this leads to

Theory binomial distribution			Simulation frequencies (10 000 bags)		
r	$H(X=r)$	$P(X=r)$	r	$H(X=r)$	$h(X=r)$
0	376	0.038	0	366	0.037
1	1352	0.135	1	1369	0.137
2	2299	0.230	2	2329	0.233
3	2452	0.245	3	2405	0.241
4	1839	0.184	4	1871	0.187
5	1030	0.103	5	1031	0.103
6	446	0.045	6	422	0.042
7	153	0.015	7	155	0.016
8	42	0.004	8	39	0.004
>8	11	0.001	>8	13	0.001

Fig. 9 Distribution of a random variable X counting the number r of red candies in 10 000 candy bags

a data-based estimation of probability, e.g. $P(\text{red}) = 1/6$ for drawing a red colored candy and $P(\text{not red}) = 5/6$ for drawing another color. Based on these assumptions, it is possible to model the situation by a binomially distributed random variable X that counts the number r of red colored candies:

$$P(X=r) = \binom{18}{r} \cdot \left(\frac{1}{6}\right)^r \cdot \left(\frac{5}{6}\right)^{18-r}.$$

Using this model, the students can calculate the different probabilities for different numbers of red candies as well as fictive absolute numbers of candy bags containing r red colored candies. Figure 9 (left) contains the probability distribution and the distribution of absolute frequencies resulting from calculating with 10 000 fictive opened candy bags. If students do not know the binomial distribution, they could use a computer simulation (prepared by the teacher). Thus, they are able to simulate 10 000 fictive candy bags and to calculate the absolute as well as the relative frequencies of different numbers of red colored candies (Fig. 9, right).

The arrays of Fig. 9 could be used to derive criteria for validating the underlying theoretical model of equal color distribution with regard to reality. For example, the students could decide to reject the model when there are 0 or 8 (or even more) red colored candies in a new real candy bag. In this case, they would be wrong by refusing the underlying theoretical model in about 5 % of cases.

This color distribution problem is structurally equivalent to the situation of throwing the cuboid die. Thus, to model the problem situation of the candy bags, the frequentistic approach of probability seems to be necessary. A theoretical model could not be convincingly found without a sample of real data which is big enough. Data are again needed to validate the theoretical model and its implications.

In comparison with the Cliffhanger Problem, the situation of the candy bags is much less artificial. Although this problem includes modelling real data, the situation contains, however, a virtual problem of the real world. In fact, the color distribution of candy bags is not a real problem if students' life or society is regarded. Nevertheless, within a pedagogical point of view, such data based problem situations represent situated learning (Vosniadou 1994). They are further to be considered worthwhile because they are anchored in the real world of students' daily life and provoke students' activity. Activity-based methods have been recognized to be a pedagogically extremely valuable approach to teaching statistical concepts (e.g. Scheaffer et al. 1997; Rossman 1997). By this, such a data based approach should encounter learning scenarios concerning probability which were criticized by Greer and Mukhopadhyay (2005, p. 314) as consisting of "exposition and routine application of a set of formulas to stereotyped problems".

5.3 A Real World Problem Situation Concerning the Subjectivist Approach

The following news item is given:
"The local health authority advises against using an HIV rapid test"

The rapid test has the following characteristics:

- If a person is infected with HIV (I), the rapid test yields a positive result ($+$) in 99 % of cases, $P(+|I) = 0.99$, but wrongly yields a negative result ($-$) in 1 % of cases, $P(-|I) = 0.01$.
- If a person is not infected with HIV ($\overline{I}$), the rapid test yields a negative result ($-$) in 98 % of cases, $P(-|\overline{I}) = 0.98$, but wrongly yields a positive result in 2 % of cases, $P(+|\overline{I}) = 0.02$.

Referring to these characteristics, the rapid test seems to be very good. So what might be the reason for the local health authority to publish this news item?

The characteristics of the rapid test represent objective frequentistic probabilities that should be based on empirical studies. However, for modelling the given situation, the situation must be regarded from a subjective perspective. By doing this, the relevant question concerns the probability of being infected with HIV given a positive test result. This situation comprises no randomness since a specific person is either infected or not infected. If this specific person has no information about his status of infection, he or she might hold an individual degree of belief, i.e. a subjective probability, of being infected or not. This subjective probability might be based on an official statement of the prevalence in a specific country, e.g. the prevalence in the Western European countries is about 0.1 %. However, this subjective probability could be higher (or lower), if a person takes into account his individual environment. Processing the test characteristics using the base rate of 0.1 % as "a priori" probability yields:

$$P(I|+) = \frac{P(+|I) \cdot P(I)}{P(+|I) \cdot P(I) + P(+|\overline{I}) \cdot P(\overline{I})} = \frac{0.99 \cdot 0.001}{0.99 \cdot 0.001 + 0.02 \cdot 0.999} \approx 0.047.$$

Although the information of a positive test result yields a bigger “a posteriori” probability concerning an individual degree of belief referring to the status of infection, the probability to actually be infected given a positive test result is very low. However, the test itself is not bad, but the test applied concerning a tiny base rate. Given a higher base rate, i.e. a subjective probability “a priori”, yields fundamentally different results. For example, the base rate of 25 %, which is an existing estimation of prevalence in some countries (www.unaids.com), results in $P(I|+) \approx 0.943$.

The structure of this real world problem is similar to the die game we discussed in Sect. 4.3. Thus, a model “a priori” (*looking ahead*) is revised by information (*looking back*), resulting in a model “a posteriori” (*looking ahead*). The difference between the rapid test and the die game is that the die could produce further information being independent from that known earlier. By contrast, it is not appropriate to repeat the rapid test to gain more than a single piece of information since the results of repeated tests could reasonably not be modelled as being independent.

As opposed to the former two problem situations, the HIV task comprises a real world problem that is relevant to society (equivalent problem situations are breast cancer and mammography, BSE, or even doping). The problem situation shows the impact of the base rate on decision making using the subjectivistic approach of probability. On the one hand, such a problem situation aims to illustrate that elementary stochastics provides methods to reconstruct problems that are relevant to life or society. On the other hand, such a problem complies with the demands of situated learning as well as anchored instruction (see above). Thus HIV infection involves an authentic and interesting real world problem situation aiming at students’ comprehension of the subjectivistic approach of probability.

6 The Role of Simulations

By reflecting on the different kinds of randomness concerning problem situations which are described above, simulations could play different roles within the underlying modelling activities. As well as, e.g. Girard (1997); Coutinho (2001); Zimmermann (2002); Engel (2002); Biehler (2003); Batanero et al. (2005a, 2005b); Pfannkuch (2005); Eichler and Vogel (2009, 2011); Biehler and Prömmel (2011), we see possible advantages within a simulation-based approach to model situations with randomness, although there are also possible disadvantages to be kept in mind (e.g. Countinho 2001). Within the problem situations discussed above, we identify three different roles, i.e. simulation as a tool to:

1. Explore a model that already exists,
2. Develop an unknown model approximately, and
3. Represent data generation.

We briefly discuss these three roles of simulation including aspects of what Mishra and Koehler (2006) call technological pedagogical content knowledge (TCPK) (with regard to probabilities, see also Heitele 1975; Biehler 1991; Pfannkuch 2005).

6.1 Simulation as a Tool to Explore an Existing Model

Concerning all problem situations we have discussed, simulation could yield relevant insights into the process of data generation within a model (Stohl and Tarr 2002). We emphasized this aspect concerning throwing the ordinary die: Simulation could illustrate possible results of relative frequencies dependent on the underlying model. In particular, simulation could (qualitatively or quantitatively) yield insight into the $\frac{1}{\sqrt{n}}$-law as well as into the empirical law of large numbers and, thus, yield "a deeper understandings of how theoretical probability [...] can be used to make inferences" (Stohl and Tarr 2002, p. 322).

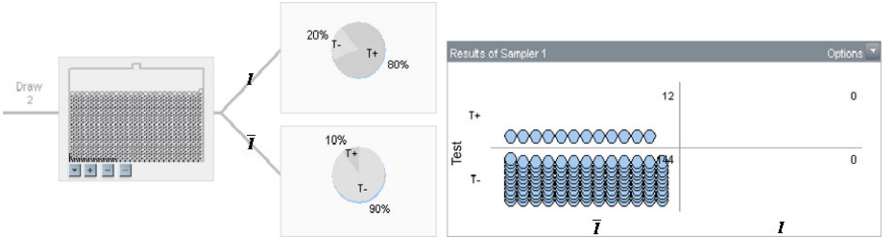
Simulations concerning a well-known model (the ordinary die) could support students when they have to make the transfer from the estimation of future frequencies based on such a well-known model to the estimation of a frequentistic probability based on relative frequencies concerning a (large) data sample. Thus, the theoretical world of probability and the empirical world of data could be convincingly connected. This is, in our opinion, one of the central pedagogical impacts of simulations: allowing for bridging these two worlds (cf. Eichler and Vogel 2011; Vogel and Eichler 2011; Batanero et al. 2005b).

6.2 Simulation as a Tool to Explore an Unknown Model

In case of the virtual problem situation of Cliffhanger, computer-based simulations producing virtual empirical data could support students by assuring them of their theoretical considerations before, during or after the modelling process. Further, simulation could be used to explore the distribution of red candies even if the (theoretical) binomial distribution is unknown. In general, simulations could help build a model where irrelevant features are disregarded, and the mathematical phenomenon is condensed in time and available to the students' work (Engel and Vogel 2004). Thus, formal mathematics could be reduced to a minimum, allowing students to explore underlying concepts and to experiment with varying set-ups. In this regard, the didactical role of simulations, which in sense of Coutinho (2001) could be considered as being *pseudo-concrete*, is to implicitly introduce a theoretical model to the student, even when mathematical formalization is not possible (cf. Henry 1997). Simulations are at the same time physical and algorithmic models of reality since they allow an intuitive work on the relevant model that facilitates later mathematical formalization (cf. Batanero et al. 2005b). Moreover, understanding of methods of inferential statistics could be facilitated by simulating a virtual but empirical world of data (cf. Pfannkuch 2005).

6.3 Simulation as a Tool for Visualizing Data Generation

In the former two sections, we described simulation as a tool to facilitate and to enhance students' learning of objective probabilities. Actually, the benefit of simulation concerning the subjectivistic approach of probability seems to be not as important as if objective probability is considered. Thus, regarding, for instance, the HIV infection problem simulation conducted with educational software like Excel, Fathom, Geogebra, Nspire CAS, and even professional software like R yields frequencies of an arbitrary number of tested persons. However, these simulation results seem to have no advantage in opposite to graphical visualizations like the tree with absolute frequencies (Wassner and Martignon 2002). By contrast, visualizing the process of data generation by simulation might facilitate the comprehension of a problem's structure concerning the subjectivistic approach (e.g. the HIV infection problem).



Educational software like Tinkerplots is suitable to visualize the process of data generation. In the box on the left side, the population according to the base rate is represented. In the middle, the characteristics of a diagnostic test are shown. On the right side, the results are represented using a two by two table (the test characteristics concern mammography and breast cancer; Hoffrage 2003). The advantage of such a simulation is to be seen in running the process of data generation in front of the learner's eyes. With regard to recent and current research in multimedia supported learning environments, there is a lot of empirical evidence (e.g. Salomon 1993; Zhang 1997; Schnotz 2002; Vogel 2006) that using them the learners can be successfully supported in building adequate mental models of the problem domain, here in learning about conditional probabilities.

7 Conclusion

There are three approaches of modelling situations with randomness: the *classical approach*, the *frequentistic approach* and the *subjectivistic approach*. With regard to a pure educational perspective, we asked for what teachers must know about the underlying concept of probability. To answer this question, we first made a short proposal of essentials of a basic content knowledge teachers should have and contrasted it with empirical findings of existing research concerning teachers' knowledge and beliefs about probability approaches. On this base, we developed a normative proposal for teaching the three approaches of probability: firstly, we outlined our normative focus on teachers' potential pedagogical content knowledge (PCK) concerning the construct of probability, and secondly, we elaborated the underlying ideas by examples of three different classroom activities of throwing dice that provide the three approaches mentioned above. Thereby we used a modelling perspective which allows for interpreting the different approaches to probability as being different kinds of modelling a problem situation of uncertainty.

This modelling perspective also allows for leaving the gambling world and for applying different approaches to probability in problem situations which could be categorized along a scale between a virtual world, on the one side, and the real world, on the other side. In a nutshell, we finally focused on technological pedagogical content knowledge (TCPK). By discussing the question how technology, and simulation in particular, can enhance students understanding of probabilities, we distinguished three different roles which simulations can play. Especially computer-based simulations are estimated to being worthwhile because of their power of producing new data which could be interpreted as empirical realizations of a random process. But by using digital tools in modelling situations with randomness, new research questions arise.

References

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching. *Journal of Teacher Education*, *59*(5), 389–407.

Batanero, C., & Sanchez, E. (2005). What is the nature of high school students' conceptions and misconceptions about probability. In G. A. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 241–266). New York: Springer.

Batanero, C., Henry, M., & Parzysz, B. (2005a). The nature of chance and probability. In G. A. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 15–37). New York: Springer.

Batanero, C., Biehler, R., Maxara, C., Engel, J., & Vogel, M. (2005b). *Using simulation to bridge teachers content and pedagogical knowledge in probability*. Paper presented at the 15th ICMI study conference: the professional education and development of teachers of mathematics. Aguas de Lindoia, Brazil.

Begg, A., & Edwards, R. (1999). Teachers' ideas about teaching statistics. In *Proceedings of the 1999 combined conference of the Australian association for research in education and the New Zealand association for research in education*. Melbourne: AARE & NZARE. Online: www.aare.edu.au/99pap/.

Biehler, R. (1991). Computers in probability education. In R. Kapadia & M. Borovcnik (Eds.), *Chance encounters. Probability in education* (pp. 169–211). Amsterdam: Kluwer Academic.

Biehler, R. (2003). Simulation als systematischer Strang im Stochastikcurriculum [Simulation as systematic strand in stochastics teaching]. In *Beiträge zum Mathematikunterricht* (pp. 109–112).

Biehler, R., & Prömmel, A. (2011). Mit Simulationen zum Wahrscheinlichkeitsbegriff [With simulations towards the concept of probability]. *PM. Praxis der Mathematik in der Schule*, *53*(39), 14–18.

Blum, W. (2002). ICMI-study 14: applications and modelling in mathematics education—discussion document. *Educational Studies in Mathematics*, *51*, 149–171.

Borovcnik, M. (1992). *Stochastik im Wechselspiel von Intuitionen und Mathematik [Stochastics in the interplay between intuitions and mathematics]*. Mannheim: BI-Wissenschaftsverlag.

Borovcnik, M. (2005). Probabilistic and statistical thinking. In *Proceedings of the fourth congress of the European society for research in mathematics education*, Sant Feliu de Guíxols, Spain (pp. 485–506).

Box, G., & Draper, N. (1987). *Empirical model-building and response surfaces*. New York: Wiley.

Brown, J. S., Collins, A., & Duguid, P. (1981). Situated cognition and the culture of learning. *Educational Researcher*, *18*(1), 32–42.

Burill, G., & Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in teacher training. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics—challenges for teaching and teacher education* (pp. 57–70). New York: Springer.

Broers, N. J. (2006). Learning goals: the primacy of statistical knowledge. In A. Rossman & B. Chance (Eds.), *Proceedings of the seventh international conference on teaching statistics*, Salvador: International Statistical Institute and International Association for Statistical Education. Online: www.stat.auckland.ac.nz/~iase/publications.

Carnell, L. J. (1997). *Characteristics of reasoning about conditional probability (preservice teachers)*. Unpublished doctoral dissertation, University of North Carolina-Greensboro.

Chaput, B., Girard, J.-C., & Henry, M. (2011). Frequentistic approach: modelling and simulation in statistics and probability teaching. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics—challenges for teaching and teacher education* (pp. 85–96). New York: Springer.

Chernoff, E. J. (2008). The state of probability measurement in mathematics education: a first approximation. *Philosophy of Mathematics Education Journal*, *23*.

Cognition Technology Group at Vanderbilt (1990). Anchored instruction and its relationship to situated cognition. *Educational Researcher*, *19*(6), 2–10.

Countinho, C. (2001). *Introduction aux situations aléatoires dés le Collége: de la modélisation à la simulation d'experiences de Bernoulli dans l'environment informatique Cabri-géomètre-II [Introduction to random situations in high school: from modelling to simulation of Bernoulli experiences with Cabri-géomètre-II]*. Unpublished doctoral dissertation, University of Grénoble, France.

De Fineti, B. (1974). *Theory of probability*. London: Wiley.

Dugdale, S. (2001). Pre-service teachers' use of computer simulation to explore probability. *Computers in the Schools*, *17*(1/2), 173–182.

Eichler, A. (2008a). Teachers' classroom practice and students' learning. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE study: teaching statistics in school mathematics. Challenges for teaching and teacher education. Proceedings of the ICMI study 18 and 2008 IASE round table conference*. Monterrey: ICMI and IASE. Online: www.stat.auckland.ac.nz/~iase/publications.

Eichler, A. (2008b). Statistics teaching in German secondary high schools. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE study: teaching statistics in school mathematics. Challenges for teaching and teacher education. Proceedings of the ICMI study 18 and 2008 IASE round table conference*. Monterrey: ICMI and IASE. Online: www.stat.auckland.ac.nz/~iase/publications.

Eichler, A. (2011). Statistics teachers and classroom practices. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics—challenges for teaching and teacher education* (pp. 175–186). New York: Springer.

Eichler, A., & Vogel, M. (2009). *Leitidee Daten und Zufall [Key idea of data and chance]*. Wiesbaden: Vieweg + Teubner.

Eichler, A., & Vogel, M. (2011). *Leitifaden Daten und Zufall [Compendium of data and chance]*. Wiesbaden: Vieweg + Teubner.

Engel, J. (2002). Activity-based statistics, computer simulation and formal mathematics. In B. Phillips (Ed.), *Developing a statistically literate society*. *Proceedings of the sixth international conference on teaching statistics (ICOTS6, July, 2002)*, Cape Town, South Africa.

Engel, J., & Vogel, M. (2004). Mathematical problem solving as modeling process. In W. Blum, P. L. Galbraith, H.-W. Henn, & M. Niss (Eds.), *ICMI-study 14: application and modeling in mathematics education* (pp. 83–88). New York: Springer.

Girard, J. C. (1997). Modélisation, simulation et experience aléatoire [Modeling, simulation and random experience]. In *Enseigner les probabilités au lycée* (pp. 73–76). Reims: Commission Inter-IREM Statistique et Probabilités.

Greer, B., & Mukhopadhyay, S. (2005). Teaching and learning the mathematization of uncertainty: historical, cultural, social and political contexts. In G. A. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 297–324). New York: Springer.

Heitele, D. (1975). An epistemological view on fundamental stochastic ideas. *Educational Studies in Mathematics*, *2*, 187–205.

Henry, M. (1997). Notion de modéle et modélization en l'enseignement [Notion of model and modelling in teaching]. In *Enseigner les probabilités au lycée* (pp. 77–84). Reims: Commission Inter-IREM.

Hoffrage, U. (2003). Risikokommunikation bei Brustkrebsfrüherkennung und Hormonersatztherapie [Risk communication in breast cancer screening and hormone replacement therapy]. *Zeitschrift für Gesundheitspsychologie*, *11*(3), 76–86.

Jones, G. A., Langrall, C. W., & Mooney, E. S. (2007). Research in probability. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 909–956). Charlotte: Information Age Publishing.

Kolmogoroff, A. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung [Fundamental terms of probability]*. Berlin: Springer. (Reprint 1973).

Mandl, H., Gruber, H., & Renkl, A. (1997). Situiertes Lernen in multimedialen Lernumgebungen [Situated learning in multimedia learning environments]. In L. J. Issing & P. Klimsa (Eds.), *Information und Lernen mit Multimedia* (pp. 166–178). Weinheim: Psychologie Verlags Union.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: a framework for teacher knowledge. *Teachers College Record*, *108*(6), 1017–1054.

Pfannkuch, M. (2005). Probability and statistical inference: how can teachers enable students to make the connection? In G. A. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 267–294). New York: Springer.

Renyi, A. (1992). *Calcul des probabilités [Probability calculus]*. Paris: Jacques Gabay. (Trans: L. Félix. Original work published 1966).

Riemer, W. (1991). *Stochastische Probleme aus elementarer Sicht [Stochastic problems from an elementary perspective]*. Mannheim: BI-Wissenschaftsverlag.

Rossman, A. (1997). *Workshop statistics*. New York: Springer.

Salomon, G. (1993). *Distributed cognitions: psychological and educational considerations*. New York: Cambridge University Press.

Scheaffer, R., Gnanadesikan, M., Watkins, A., & Witmer, J. (1997). *Activity-based statistics*. New York: Springer.

Schneider, I. (1988). *Die Entwicklung der Wahrscheinlichkeitstheorie von den Anfängen bis 1933 [Development of probability from beginnings to 1933]*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Schnotz, W. (2002). Enabling, facilitating, and inhibiting effects in learning from animated pictures. In R. Ploetzner (Ed.), *Online-proceedings of the international workshop on dynamic visualizations and learning* (pp. 1–9). Tübingen: Knowledge Media Research Center.

Schupp, H. (1982). Zum Verhältnis statistischer und wahrscheinlichkeitstheoretischer Komponenten im Stochastik-Unterricht der Sekundarstufe I [Concerning the relation between statistical and probabilistic components of middle school stochastics education]. *Journal für Mathematik-Didaktik*, *3*(3/4), 207–226.

Schupp, H. (1988). Anwendungsorientierter Mathematikunterricht in der Sekundarstufe I zwischen Tradition und neuen Impulsen [Application-oriented mathematics lessons in the lower secondary between tradition and new impulses]. *Der Mathematikunterricht*, *34*(6), 5–16.

Shaughnessy, M. (1992). Research in probability and statistics. In D. A. Grouws (Ed.), *Handbook of research in mathematics teaching and learning* (pp. 465–494). New York: Macmillan.

Sill, H.-D. (1993). Zum Zufallsbegriff in der stochastischen Allgemeinbildung [About the concepts of chance and randomness in statistics education]. *ZDM. Zentralblatt für Didaktik der Mathematik*, *2*, 84–88.

Stachowiak, H. (1973). *Allgemeine Modelltheorie [General model theory]*. Berlin: Springer.

Steinbring, H. (1991). The theoretical nature of probability in the classroom. In R. Kapadia & M. Borovcnik (Eds.), *Chance encounters. Probability in education* (pp. 135–168). Amsterdam: Kluwer Academic.

Steinbring, H. & Strässer, R. (Hrsg.) (1981). Rezensionen von Stochastik-Lehrbüchern [Reviews of stochastic textbooks]. *ZDM. Zentralblatt für Didaktik der Mathematik*, *13*, 236–286.

Stohl, H. (2005). Probability in teacher education and development. In G. A. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 345–366). New York: Springer.

Stohl, H., & Tarr, J. E. (2002). Developing notions of inference using probability simulation tools. *The Journal of Mathematical Behavior*, *21*, 319–337.

Vogel, M. (2006). *Mathematisieren funktionaler Zusammenhänge mit multimediabasierter Supplantation [Mathematization of functional relations using multimedia supplantation].* Hildesheim: Franzbecker.

Vogel, M., & Eichler, A. (2011). Das kann doch kein Zufall sein! Wahrscheinlichkeitsmuster in Daten finden [It cannot be chance! Finding patterns of probability in data]. *PM. Praxis der Mathematik in der Schule*, *53*(39), 2–8.

von Mises, R. (1952). *Wahrscheinlichkeit, Statistik und Wahrheit [Probability, statistics and truth].* Wien: Springer.

Vosniadou, S. (1994). From cognitive theory to educational technology. In S. Vosniadou, E. D. Corte, & H. Mandl (Eds.), *Technology-based learning environments* (pp. 11–18). Berlin: Springer.

Wassner, C., & Martignon, L. (2002). Teaching decision making and statistical thinking with natural frequencies. In B. Phillips (Ed.), *Developing a statistically literate society. Proceedings of the sixth international conference on teaching statistics* (ICOTS6, July, 2002). Cape Town, South Africa.

Wickmann, D. (1990). *Bayes-Statistik [Bayesian statistics].* Mannheim: BI-Wissenschaftsverlag.

Wild, C., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(3), 223–248.

Zhang, J. (1997). The nature of external representations in problem solving. *Cognitive Science*, *21*(2), 179–217.

Zimmermann, G. M. (2002). *Students reasoning about probability simulation during instruction*. Unpublished doctoral dissertation, Illinois State University, Normal.

A Modelling Perspective on Probability

Maxine Pfannkuch and Ilze Ziedins

Abstract In this chapter, we argue for three interconnected ways of thinking about probability—"true" probability, model probability, and empirical probability—and for attention to notions of "good", "poor" and "no" model. We illustrate these ways of thinking from the simple situation of throwing a die to the more complex situation of modelling bed numbers in an intensive care unit, which applied probabilists might consider. We then propose a reference framework for the purpose of thinking about the teaching and learning of probability from a modelling perspective and demonstrate with examples the thinking underpinning the framework. Against this framework we analyse a theory-driven and a data-driven learning approach to probability modelling used by two research groups in the probability education field. The implications of our analysis of these research groups' approach to learning probability and of our framework and ways of thinking about probability for teaching are discussed.

1 Introduction

The teaching of probability in schools has evolved over the last 30 years to incorporate two interpretations of probability: a theoretical notion of a mathematically computed probability distribution called the classical approach; and an empirical notion of a stabilized probability distribution when some random experiment is repeated many times under the same conditions called the frequentist approach. The consequence for students is that these two conceptions of probability, the model and the reality, may become confused and mixed up (Chaput et al. 2011). Furthermore, research has documented many misconceptions associated with probabilistic thinking (e.g. Kahneman et al. 1982; Konold 1989; Lecoutre 1992; Nickerson 2004). Also current instruction and assessment promote an impoverished representation of probability through presenting examples that unrealistically assume independence, discount context as a factor, have one right answer, and do not address alternative

M. Pfannkuch (✉) · I. Ziedins
The University of Auckland, Auckland, New Zealand
e-mail: m.pfannkuch@auckland.ac.nz

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_5,

assumptions, which can lead to different models and answers (Greer and Mukhopadhyay 2005). Often teaching regresses to a list of formulas and routine applications and as Borovcnik (2011, p. 81) observed, "probability is signified by a peculiar kind of *thinking*, which is not preserved in its mathematical conception."

Probability instruction is the one area in the school curriculum where students are afforded the opportunity to view the world non-deterministically and to learn about randomness and decision-making under uncertainty. The challenge for educators is to determine the conception of probability that should be promoted in teaching and learning. Currently, the best direction in which to steer probability learning seems to be towards a modelling approach (Chaput et al. 2011; Greer and Mukhopadhyay 2005; Konold and Kazak 2008; Shaughnessy 2003). As Jones (2005, p. 6) commented:

> It is only in the last four centuries that mathematicians have developed quantitative measures for randomness, that in turn gave birth to various conceptions of probability: the classical approach based on combinatorics, the experimental approach leading to the limit of stabilized frequencies, and finally to a conflict between objective and subjective points of view. In more recent times, we see the emergence of modelling approaches that endeavour to distinguish between real random situations and their theoretical interpretations.

Indeed, Batanero et al. (2005, p. 32) think that "interpreting random situations in terms of probabilistic models will serve to overcome the controversy between classic, subjective and frequentist approaches." From a teaching perspective, Chaput et al. (2011) describe three stages in the modelling process: pseudo-concrete model (putting empirical observations into a working model), mathematization and formalization (translating working hypotheses into model hypotheses to design probability model), and validation and interpretation in context (checking fit of a probability model to data). The question is how such a perspective can be put into practice and what would be desired learning outcomes across the curricular levels.

We believe that this new and emergent direction in the teaching of probability needs to be examined. Therefore, we will reflect on the practice and thinking of probabilists, from which we will propose a reference framework for curricular activities and learning. Using this framework, we will analyse the modelling approaches taken by two research groups. The implications of taking a modelling perspective on probability will be discussed in terms of the framework, school curricula and research into the development of students' probabilistic thinking.

2 Probability from a Modelling Perspective

Students often learn about probability primarily as a mathematical construct, with associated rules for the analysis of that construct. However, it is equally important for students to understand why and how probability models are developed and the assumptions that underpin their use. In this section, we propose a framework in which probability models can be developed and presented to students through first reflecting on the thinking that probabilists use for modelling.

2.1 Probability and Probability Models

The term "probability model" can be understood in various ways. We are thinking of a model that mimics some aspect of random behaviour in the real world. This may range from something as simple as the toss of a coin or a die, to something as complicated as modelling patient flow through a hospital unit, examples of which we will discuss in some detail. In our discussion, a probability model will often be associated with the idea of a system evolving dynamically over time, as happens, for instance, with the number of patients in a hospital intensive care unit. Such a model can be used in many ways:

1. To estimate the number of beds that are needed to accommodate the flow of patients, to better match staffing of beds to patient flow,
2. To use as a tool in considering the effect of changing some aspect of patient care or admitting additional streams of patients to the intensive care unit, and,
3. More generally, to better understand the most effective means of reducing patient waiting times for treatment and increasing quality of care.

As illustrated by this example, a model is usually built to answer a particular question or questions about a system, sometimes just to understand its behaviour better, but often in order to optimize some measure of its performance, or alternatively, to predict performance under some alternative scenario. We think that in order to motivate the development of such models, it is important that students understand the purpose of modelling and how such models may be used, and that although they are only ever approximations to what happens in the real world, these approximations can help us to better understand the behaviour of real world systems and perhaps subsequently improve their performance. In this respect, a model becomes a "virtual world" in which experiments can be conducted to ascertain the effects in the real world, ideas that are common to applied mathematics.

Before outlining a reference framework for probability modelling, we discuss our understanding of the probability of an event. There are at least three ways of thinking about the probability of an event or the probabilities of a collection of events (see Konold et al. 2011). The first is the (usually unknown) "true" probability that the event occurs. The second is the probability assigned under a given model—we refer to this as the model probability (also known as the theoretical probability). And the third is the empirical probability, which is also obtained with reference to some underlying assumptions, which can be thought of as an underlying model. Both the model probability and the empirical probability are *estimates* of the true probability (Fig. 1).

We illustrate these three ways of thinking about probability with the simple example of a coin toss. An obvious model for the toss of a coin is that it is equally likely to show heads or tails, that is, it is a fair coin. So *a* model probability for heads is $1/2$ (and often this will be an excellent model, fit for most purposes). However, in practice, a particular coin may be slightly biased, so that its true probability is not exactly $1/2$. Then, under the assumptions that each toss of the coin is independent of the others, and that each time the coin is tossed it has the same probability

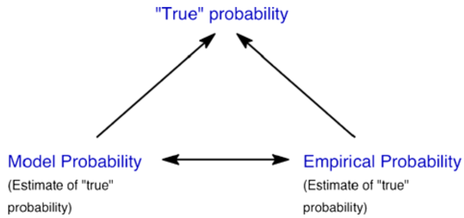


Fig. 1 Three interconnected ways to think about probability

of showing heads, the probability that the coin shows heads can be estimated by the empirical probability, the proportion of heads in a large number of tosses.

This simple example illustrates the point that rather than claiming to give the exact probability that something happens in real life, model probabilities are estimates of the true probability and arise from a model that we hope captures the essential behaviour of the process of interest to us. Model probabilities need to satisfy certain consistency conditions to be useful for probabilists and mathematicians.

With this background in mind, we propose a description of a particular situation or process as having a "good model", a "poor model", or "no model". For instance, the standard theoretical model for a fair coin toss is that heads and tails are equally likely with probability $1/2$ each, as outlined above. Repeated tosses of a fair coin can be used to estimate the probabilities of heads and tails. For a fair coin we would expect these empirical probabilities to be close to the theoretical model probabilities of $1/2$, and for this to be a good model. If we were spinning a coin, rather than tossing it, we might think that the obvious theoretical model was still that heads and tails are equally likely, but estimates of the outcome probabilities from sufficiently large experiments show that this is a surprisingly poor model (Konold et al. 2011). There is then a need to find a better model using the estimates from experiments. Finally, there may be situations where we have no model, whether good or bad, and must estimate probabilities and probability distributions via empirical probabilities. These estimates can then be used as the basis for a theoretical model. Note the adequacy of a "good" model is usually determined by experimental testing.

A "good" model means "fit for purpose". Thus a model may be perfectly adequate in one context, but may fail in another. For instance, a model may model inflation over a period very successfully, but fail to do so when some variable not included in the model changes its value; or we may have a model predicting agricultural outputs for the year, which might be adequate in the short-term but might not be in the longer term if no allowance has been made for climate change.

To summarize, the three situations that we are considering are:

- "good model"—meaning fit for purpose;
- "poor model"—meaning not fit for purpose and requiring further development;
- "no model"—meaning in this situation, we may make some underlying assumptions about a process which will permit us to develop a model.

2.2 *Three Examples*

We now explore further the ways of thinking about probability (Fig. 1) and the three model situations via some examples.

2.2.1 Throwing a Single "Fair" Die

Throwing a fair die is essentially the same as the coin tossing example, except that there are now six possible outcomes from a throw, and it seems obvious to most people that for a fair die each of the six faces are equally likely to be thrown. This is an initial model probability. However, in reality, a particular die may have a little bit chipped off one corner, or one edge might be ever so slightly skew, so that the "true" probability of throwing a two might be very close to $1/6$, but not exactly $1/6$. The difference here may be so slight that the number of throws required to detect it would be dauntingly large, and it is therefore not interesting practically. This would then be a "good model".

However, if the difference between the model probability of $1/6$, and the true probability is sufficiently large, then it will be necessary to estimate the probability of throwing a two empirically, by taking the proportion of twos seen in a large number of trials to give an improved estimate of the probability of throwing a two, and hence a better model. Notice that when we do this we usually assume that the sequence of trials is independent and identically distributed—this is one of the model assumptions. If there is some mechanism in play, which means that this is not the case, then we might need to add further elements to the model to ensure that it is fit for purpose (e.g. if a talented individual is trying to train herself to throw a two on demand, and becoming more successful at it).

One of the reasons that a "fair" die is a good example to start with is that there is such an obvious "good" model, provided we make the assumption that throws are independent and identically distributed. But this very feature can also be a drawback since it obscures the distinction between the model probability and the true probability.

2.2.2 Throwing a Basketball Through a Hoop

A basketball player is practicing shooting goals. An initial, simple model might be that the player gets the goal with probability 0.8 on every trial, independently of all other trials. However, the player might point out, very reasonably, that in a sequence of shots in a single practice session, the first few practice shots are less likely to be successful, then there is a gradual improvement, but eventually the player starts to tire and accuracy degrades. In an actual match, players may be affected by nerves and have an "off" day, or, on the other hand, a particularly good day with excellent performance. For all these reasons, the proposed model might be a "poor" model.

In this situation, the idea of a long sequence of independent and identically distributed trials giving estimates that converge to the "true" probability does not match the reality of what actually happens. To build a "better" model, it may be necessary to build a model that incorporates the increase and then decline of the success probability over a longer sequence. But how would we estimate the success probabilities for such a model? An initial attempt at an improved model might be to observe a number of sequence of shots at the same time every day, say at the beginning of the day, where the daily sequence is now a single trial. But in reality again, a player may improve over a season, so that there is no possibility of observing successive sequences under the same conditions, and a yet better model would incorporate this increasing trend.

This is an example where the true probabilities change over time (are not time-homogeneous) in such a way that they may never be accessible via finite estimates. However, it may be the case that a simpler model is adequate, if there is insufficient data either to reject it, or to fit a more complex model. For instance, if it is clear that the sequence of trials is not independent and identically distributed, we could try to first fit a Markov model (where the probability of a success depends on whether the previous trial was a success or failure), and if that is not adequate a higher order Markov model (where the probability of success depends on the outcomes of the previous two or more trials). The transition probabilities for a Markov chain can also be non-time-homogeneous, and it might be possible to fit a simple model where within a practice session the success probabilities increase and then decrease over time.

2.2.3 Modelling the Number of Beds in an Intensive Care Unit (ICU)

Modelling bed numbers in an ICU is the kind of "real" problem that applied probabilists might consider (see Chen et al. 2011). The number of occupied beds in an intensive care unit depends on both the arrival and service processes. The arrival process can be thought of as separate streams of arrivals for different patient types. Elective patients arrive at pre-determined times for their scheduled surgery. Emergency patients, on the other hand, arrive at times that are not pre-determined, and their arrival times are usually modelled as a Poisson process. The Poisson process is often a very good model for arrivals from a large population where the probability that any given individual arrives during a short time interval is very small (the Poisson approximation to the Binomial with very low success probabilities). However, the rate at which patients arrive varies with the time of day (fewer at night), day of the week (fewer at the weekend), and the season of the year (this may be due to seasonal incidence of conditions that are treated at the ICU, but also due to, for instance, lower rates of electives being treated during holiday seasons). Note how the creation of a model is inextricably linked to the data.

The simplest model might assume that emergency arrivals occur as a Poisson process with constant rate per day, which is estimated in the usual fashion by taking the total number of arrivals divided by the number of days over which the data has been gathered. If we only wish to model the number of average arrivals per day to the unit over a given time period, then this may be a sufficiently good model. However, we may be interested in answering questions such as the following:

- How many beds need to be staffed so that the probability an emergency patient does not find a bed free is less than, say, 0.01?
- How many additional beds would be needed if the overflow of patients, who cannot be admitted to an adjacent ICU, were directed here?
- How will staffing and capacity requirements be changed if a new discharge process is introduced which results in patients staying in the ICU for an hour less, on average and, by implication, is it worth trying to change the discharge process to achieve this?
- How many additional beds are needed if we wish to reduce the number of cancellations of elective surgery by, say, 50 %?

All of these questions require a more complex model. A more complex model might still assume that arrivals are as a Poisson process, but with a time-varying rate. A first estimate of these rates may be made by assuming that successive weeks have the same rates, and are independent of each other. For instance, the rate per hour at which arrivals occur from 12:00 to 16:00 on Mondays could be estimated by taking the total number of arrivals observed during that period over successive weeks, and dividing by the total number of hours observed. But the underlying conditions may vary not only with the time of day, the day of week and the season. On a longer time scale the population size changes, the demographics change (with say an ageing population), and these will also affect the rate of arrivals at the intensive care unit. If we are interested in arrival rates on these longer time scales, then we need to allow for these changes also in the estimates.

In practice, the model that is built will be the simplest that is fit for purpose, that is, that can answer the questions that have been posed about the process, which we are modelling. Another way of wording this is that the purpose of the model and the questions that will be asked of it determine its complexity. The validation process for the model often includes a testing stage where elements that have been omitted from the model are introduced to see whether they alter any of the conclusions that have been drawn.

The true arrival process in this ICU situation is changing with time and is unknown and there does not seem to be a simple estimate of an underlying rate as there was for the throw of a die. Hence, we are forced to consider a more complex notion of an estimated rate. Crucially, these estimated arrival rates are driven by the model that we are trying to fit to the data. The model and its associated parameter estimates then give estimates of probabilities for events of interest. The parameters that we try to estimate are model-dependent. Even for the simplest case, the throw of a single die, we are making some assumptions about the model. At the very least we usually assume that successive throws are independent, and have the same distribution, even if the probabilities of each face being thrown are not exactly $1/6$. In the ICU example, we see how the estimate of the arrival rate can be successively refined, with reference to the data, as our model becomes more complex.

2.3 A Proposed Reference Framework for Developing a Probability Modelling Perspective

Our proposed framework (Fig. 2), derived from thinking about probability from a modelling perspective, incorporates three components—purpose, theory-driven and data-driven. The purpose component, discussed in Sect. 2.1, has been split into two to show explicitly the process used whereas in reality the two parts are linked. In the previous section, we developed the idea that both model probabilities and empirical probabilities are estimates of a "true" probability. These two kinds of probabilities encapsulate, in simple form, a theory-driven and data-driven approach to probability and the links between them, on which we will now elaborate.

Both the theory-driven and data-driven approaches start with prior knowledge, which depends on the level of the learner. No matter what the level of the learner is, all learners bring their understandings or "theories", based on their contextual knowledge, experiences or intuitions, to probabilistic situations. At the middle school level, Konold and Kazak (2008) describe their approach to students learning about the sum of two dice. Initially, the students are asked to select from five possible distributions for the sum of two dice the distribution they would most likely get if they tossed two dice 1000 times and plotted the sums. Most students choose a distribution where the sums from four to 10 are equally likely, with the other sums less likely. Although viewed pedagogically by Konold and Kazak as a way of engaging students in their learning, from our framework perspective this activity is actually assisting the students to think probabilistically by encouraging them to adopt an initial theoretical view of the real world situation before embarking on creating a model.

After their students have chosen an initial model they move to the data-driven component of the framework where they generate data through a hands-on activity followed by a computer simulation of throwing two dice and plotting the sum. Drawing on the strategies and tools of the law of large numbers, and notions of signal and noise, and distribution, students fit a probability model to the data, and consequently revise their initial theoretical view of the real world system (Konold and Kazak 2008). Finally, with a working model in hand they return to the theory-driven component, where the "good" model is now re-interpreted as the theoretical model. For this particular example, however, students could remain entirely within the theory-driven component and mathematically or systematically work out the number of possible ways to obtain each outcome. But such an approach would deprive students of an opportunity to gain insight into the relationship between data and models.

The ideas that naturally emerge as fundamental when considering such models are ideas of randomness, independence, distributions, and sequences of independent and identically distributed trials. These are some of the basic probability concepts. Students may have good intuitions about randomness (Pratt 2005) but ideas about random behaviour which are connected to independence and distributional notions are difficult for students to grasp (Watson 2005). For example, a very commonly expressed belief is that if a long sequence of heads has been observed then the

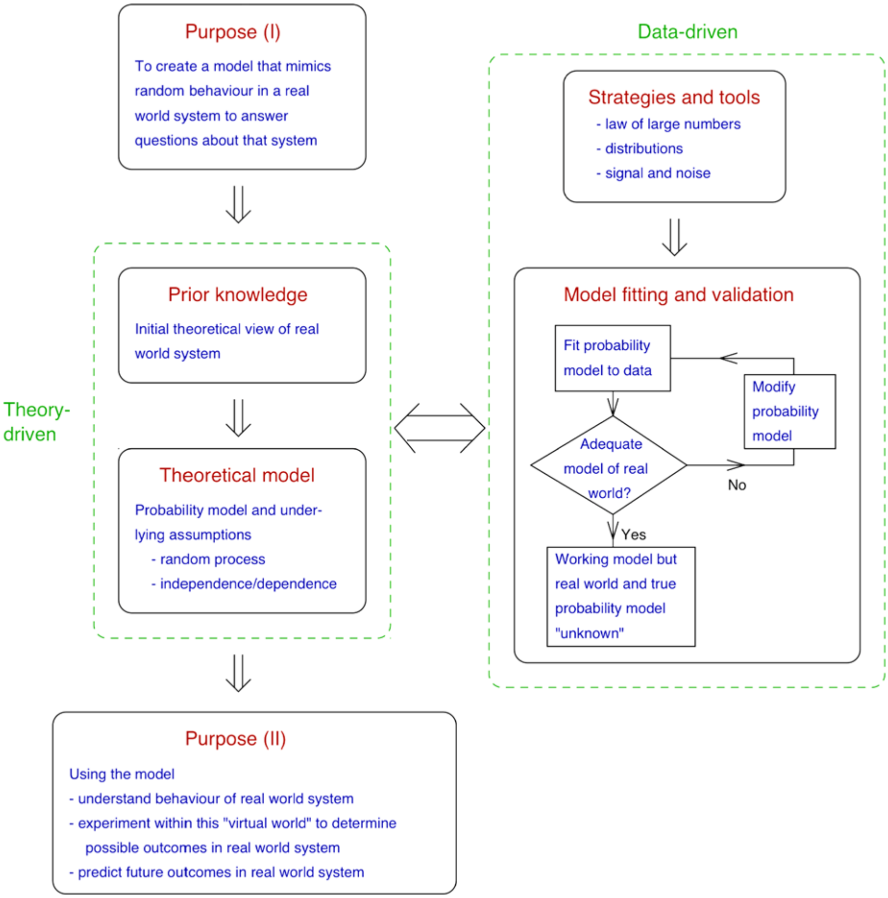


Fig. 2 Framework to think about probability from a modelling perspective

next throw is more likely to be a tail. Thus the idea of randomness linked with independence can be challenging for students.

As an aside, estimating probabilities may require a conceptual leap for students. Estimating the height of a student via repeated measurements is a relatively straightforward task—there is a physical entity to be measured repeatedly, and then an average is taken of those measurements. However, when estimating a probability (such as the probability of throwing heads) there is no physical entity to be measured, each outcome is either heads or tails, and the number of observations determines the granularity of the estimate. Thus if a coin is tossed twice, the estimate will be either 0, $1/2$, or 1, whereas when measuring a physical entity, the granularity of the estimate is determined by the accuracy of the measure used. Similarly, if we have a distribution for someone's height, then that gives us some concept of what the range of the next measurement will be, whereas while knowing that a coin is fair tells us that there is an equal chance of observing heads and tails at the next throw, the outcome will nevertheless always be one of heads or tails, regardless of how fair or not the coin is. The idea of a probability and its estimate is therefore inextricably linked with a sequence of trials, a process observed over a period of time, and the law of large numbers.

Considering the theory-driven component again, it is possible to start with some initial assumptions about the behaviour of a process and remain completely within the theory-driven component without ever moving to the data-driven component. For instance, suppose we have a situation based on a single-server queue. At the tertiary student level prior knowledge might include building blocks such as the Poisson process and distributions such as the exponential distribution. One of the standard queuing models for a single-server queue assumes that arrivals to the queue are a Poisson process, that customers require exponentially distributed service times, and that all inter-arrival times and service times are independent of one another. Then questions that are standardly asked, and answered, for such a system are the following: What is the probability, in equilibrium, that there are n customers in the queue? What is the expected number of customers in the queue? What is the expected length of an idle period for the server? What is the proportion of time that the server is busy? How long does an arriving customer need to wait for service, in equilibrium?

There are very strong links between the data-driven and theory-driven approaches. Theoretical models have often (maybe even always) arisen as an abstraction of some aspect of a real world process, although their study may then continue long afterwards without further reference to such a motivation since the mathematical models are of interest in their own right. The models that we choose to fit to data often satisfy distributional assumptions and assumptions of independence that then make it easier to work with the theoretical model. For both approaches, the ways in which we seek to generalize models or break assumptions made in models may be suggested by the other approach. Thus modifications to models that are of interest in practice may suggest further theoretical work, and more general theoretical models that are already readily available may suggest the next model to fit when seeking to improve the fit of a model to data.

3 Some Probability Modelling Learning Approaches

In Sect. 2, we discussed how probability from a modelling perspective operated in three examples from which we proposed a reference framework for the purpose of thinking about the teaching and learning of probability. Reflecting on these three examples for the purposes of student learning, many questions emerge about how to enculturate students into this type of probability practice and thinking. Questions such as:

- What elements in practice and thinking are fundamental to developing students' conceptions of probability modelling?

- What conceptual pathway across the school curriculum and beyond will lead students to learn about and appreciate the power of modelling?
- What learning experiences will assist students to understand and use probability models to answer questions of interest to them?
- How can interesting and engaging questions, such as the ICU example, be translated into a classroom activity?

There are many research groups working in the probability education field such as Pratt (2005), Stohl and Tarr (2002), Batanero and Sanchez (2005) and Shaughnessy (2003). However, in order to think about these questions we have chosen two particular research groupings that illustrate distinct approaches to teaching probability modelling. Therefore, we discuss and analyse against our framework the Borovcnik theory-driven approach to teaching probability modelling to senior school students (Borovcnik 2006; Borovcnik and Kapadia 2011), the Konold data-driven approach to middle school students (Konold et al. 2007; Konold and Kazak 2008; Lehrer et al. 2007), and a combined data and theory-driven approach.

3.1 Theory-Driven Approach

The Borovcnik approach is theory-driven, not data-driven. The problem and "data" are given, from which an underlying model is recognized and proposed such as the Binomial, Poisson or normal distribution. The model is assumed to be "good" and therefore is not built or tested. Students work within the closed world of the model; they ask questions of the model, look at the consequences of the model, and use the model to choose between several actions to improve, for example, the cost of a process. That is, *Purpose* (*II*) in our framework is realized as students use their model to predict future outcomes in the real world system.

The discussion on how to improve the model centres on multiple interpretations of a given parameter and the consequential solutions. Borovcnik and Kapadia (2011) present "the Nowitzki Task" involving success rates in free basketball throws, where the Binomial assumptions of independent trials and the probability of a success, p, being the same for each trial, are problematic. For this example, however, they do not discuss the resolution of these problems but they do demonstrate how the solution for the probability of scoring exactly 8 points in 10 trials could change through using different interpretations of p, namely "(i) p is known, (ii) p is estimated from the data [in which case a confidence interval for the probability p is considered], (iii) p is hypothesized from similar situations" (Borovcnik and Kapadia 2011, p. 15). From our perspective they still work within the theory-driven model but look at the consequences and validity of different interpretations of p, resulting in the realization that there is not one answer to the problem. According to Greer and Mukhopadhyay (2005), this type of learning experience is essential for students to develop non-deterministic ideas and to appreciate the modelling process.

The types of problems presented in the Borovcnik learning approach are important for recognizing and using particular probability models and their underlying assumptions, properties, and distributions. The Borovcnik approach, however, may lead to students viewing probability entirely as a mathematical or theoretical exercise, rather than having an image of probability as a repeatable process from an empirical perspective. Only the latter view, according to Liu and Thompson (2007), provides real evidence of development of a probabilistic way of thinking. Furthermore, in our proposed framework, theory-driven analysis is only one part of the modelling approach and hence for students to gain a better appreciation of probability modelling they need data-driven experiences as well.

3.2 Data-Driven Approach

The Konold approach is essentially data-driven. We will first discuss simple probability situations and then more complex situations where multiple factors are modelled.

3.2.1 Simple Probability Situations

Konold et al. (2011) suggest that students exposed entirely to probability situations that could be ascertained theoretically and empirically, such as the sum of two dice, coin tossing, and number of boys in a four-child family, could develop an impoverished view of probability. Hence they believe that students should experience problems for which there is no theoretical probability model or the presumed theoretical model is inadequate. For example, for the game "pass the pig" there is no theoretical model for the outcomes of tossing a pig. By physically inspecting the pig, an initial theoretical viewpoint could ascertain which outcomes were more likely than others. With the underlying assumptions that the tossing is a random process, and each trial is independent, a good model can be developed through tossing the pig many times to obtain an empirical distribution of outcomes, which is able to predict future outcomes. Similarly, rolling a hexagonal pencil might start with a theory-driven model of equiprobable outcomes. Through rolling a hexagonal pencil many times, the theory-driven model can be found to be an inadequate or poor model, and hence the empirical model of outcomes becomes the good model, and consequently the new theoretical probability model. In terms of the framework, both the theory-driven and data-driven parts are covered, but the students do not experience actually modifying the probability model, only modifying their belief about their initial theoretical model, as all that is required is a sufficiently large number of trials to obtain a working model. The purpose of the modelling for these two examples can be realized if set in game scenarios.

3.2.2 More Complex Probability Situations

The Konold teaching approach introduces middle school students to probability modelling through measurement activities, such as obtaining multiple measures of one person's head circumference. Such a hands-on activity allows students to (i) get a sense of the distribution of the measurement data for one head circumference, which is collated from the class and (ii) appreciate the types of measurement errors that can occur. The students in their studies identified sources or causes of measurement errors such as rounding, reading, slippage of tape, and angle perception (Konold and Kazak 2008; Lehrer et al. 2007). Drawing on both these learning experiences, the students used chance simulations built into TinkerPlots (Konold and Miller 2005). First, students built a model with each "model observed measurement" represented by the sum of the random errors (where each random error was calculated from a spinner representing the source of the error) and the median of the actual observed measurement, which represented an estimate of the true head circumference. Note that the building of the model required the students to have experience and insight into the real world situation, that is, an initial theoretical view of the real world system for their probability model. Second, the students ran the simulations many times and then compared the distribution produced by the model with the actual data distribution. Third, by recognizing that their model was inadequate or poor, students redesigned some of their spinners. Fourth, they ran the simulations again and considered that they now had a good model, as it was "a better match to the shape and centre of the observed values" (Lehrer et al. 2007, p. 212).

Underlying these learning experiences of a data-driven analysis but not made explicit to the students were the assumptions of a random process, and the independence of each "model observed measurement." According to our framework, the Konold modelling approach to probability is eliciting probability practice and thinking for the data-driven part of the framework, as students fit a probability model to data, realize the model is an inadequate model of the real world, and then adjust the model until they obtain a working model that adequately reflects actual data. Students also learnt how to produce different distributions by manipulating properties of the model. For students the challenge in the activity is creating a virtual world that seems to match the real world. But once this virtual world is created, what should come next? The purpose of modelling is to predict future events or to experiment in the virtual world to find out the consequences of what might happen in the real world if a certain course of action is taken. Hence the challenge for educators appears to be thinking of "what if" and prediction questions once the virtual world is created to enable students to appreciate fully the usefulness, power, and purpose of probability modelling.

Another activity that the Konold group devised was for students to create population models for units such as cats, skateboards, and candies (Konold et al. 2007). Based on their contextual knowledge of the probability distributions of attributes, the students created probability models, which they matched against their imagined population distribution of the attribute, to determine whether they had obtained a good model of the situation. The students seem to be learning that models can depict reality. Again the purpose of modelling was not in the foreground. Much more research and innovative work needs to be conducted in this area to find ways into opening up the richness of probability modelling to students.

The Borovcnik and Konold approaches illustrate two different ways of thinking about learning probability from a modelling perspective. From the viewpoint of our framework, we would argue that students need both approaches including the probability situations addressed in the next section.

3.3 Combined Theory and Data-Driven Approach

For some probability situations such as the sum of two dice, building a probability model is possible empirically and theoretically. Since these situations are typically in the school curriculum, a lot of research has been conducted in this area (e.g. Abrahamson 2006; Ireland and Watson 2009; Konold and Kazak 2008; Lee and Lee 2009). In many of these situations, students are unable to experience ideas about good and bad models; rather they become aware that probabilities and probability distributions can be obtained theoretically or empirically. Analogous to the simple probability situations in Sect. 3.2.1, the students do not actually modify their probability models; only conduct a sufficiently large number of trials to determine that the empirical and theoretical probability distributions are similar. The problem with only exposing students to these situations is that they may develop misconceptions and may not realize or appreciate the rationale for using the two approaches and the connections between them (Chaput et al. 2011; Konold et al. 2011). Again the purpose of modelling, without well-developed task scenarios, may remain opaque to many students.

4 Conclusion

Teaching and learning probability from a modelling perspective is only now possible because the technology and tools are available. Also the technology encroaching on everyday life has opened up people's minds to the possibility of operating in virtual worlds. Hence the time is ripe to consider new ways of teaching probability. Tertiary courses endeavour to move students to a world where they work with ever richer and more complex models and seamlessly move between real and virtual world systems. School students are already experiencing some of the elements of theory-driven and data-driven approaches. If they are to experience modelling, they may need to start by exploring the notion of a model as an approximation to reality, which may be fit for some purposes, but not others, and which can be improved if necessary. Teaching from a modelling perspective, moreover, would better resemble the practice of applied probabilists and should help students build a more integrated view of probability and a probabilistic way of thinking. We believe that if teaching reframed probability in terms of our triangular process for ways of thinking about probability (Fig. 1), attended to notions of "good", "poor", and "no" model, and took cognizance of our framework (Fig. 2), the typical activity of throwing a die might begin to make more sense to students.

Against our framework (Fig. 2), current attempts to teach probability from a modelling perspective suggest that researchers are pursuing interesting and far-reaching directions. The necessary student learning experiences appear to be:

1. Appreciating the circularity between theory-driven and data-driven probability modelling;
2. Cognitively integrating prior theories, known models, assumptions (e.g. random process, independence/dependence), and conceptual strategies (e.g. fitting models, law of large numbers);
3. Realizing and experiencing the purpose of modelling; and
4. Experiencing all three interdependent approaches (theory-driven, data-driven, combined theory and data-driven).

To coordinate all these aspects into a workable pathway through the school curriculum will be a challenge to educators. We believe, however, that our framework represents current probability practice and therefore sets the compass for educators to seek out the conceptual foundations of that practice and to provide researched learning trajectories that will enculturate and develop in students the propensity to think and act from a probability modelling perspective.

References

Abrahamson, D. (2006). Learning chance: lessons form a learning-axes and bridging tools perspective. In A. Rossman & B. Chance (Eds.), *Proceedings of the seventh international conference on teaching statistics*, Working cooperatively in statistics education, Salvador, Brazil, July 2006, Voorburg: International Statistical Institute. Online: http://www.stat.auckland.ac.nz/~iase/publications.php.

Batanero, C., & Sanchez, E. (2005). What is the nature of high school students' conceptions and misconceptions about probability? In G. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 241–266). New York: Springer.

Batanero, C., Henry, M., & Parzysz, B. (2005). The nature of chance and probability. In G. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 15–37). New York: Springer.

Borovcnik, M. (2006). Probabilistic and statistical thinking. In M. Bosch (Ed.), *Proceedings of the fourth congress of the European Society for Research in Mathematics Education*, Sant Feliu de Guixols, Spain, 17–21 February 2005 (pp. 484–506). http://ermeweb.free.fr/CERME4/.

Borovcnik, M. (2011). Strengthening the role of probability within statistics curricula. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics—challenges for teaching and teacher education: a joint ICMI/IASE study: the 18th ICMI study* (pp. 71–83). New York: Springer.

Borovcnik, M., & Kapadia, R. (2011). Modelling in probability and statistics. In J. Maasz & J. O'Donoghue (Eds.), *Real-world problems for secondary school mathematics students: case studies* (pp. 1–43). Rotterdam: Sense Publishers.

Chaput, B., Girard, J., & Henry, M. (2011). Frequentist approach: modelling and simulation in statistics and probability teaching. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics—challenges for teaching and teacher education: a joint ICMI/IASE study: the 18th ICMI study* (pp. 85–95). New York: Springer.

Chen, W., Frew, K., Ihaka, R., McKee, A., & Ziedins, I. (2011). Simulation and capacity planning for a cardio-vascular intensive care unit (in preparation).

Greer, B., & Mukhopadhyay, S. (2005). Teaching and learning the mathematization of uncertainty: historical, cultural, social, and political contexts. In G. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 297–324). New York: Springer.

Ireland, S., & Watson, J. (2009). Building a connection between experimental and theoretical aspects of probability. *International Electronic Journal of Mathematics Education*, *4*(3), 339–370.

Jones, G. (2005). Introduction. In G. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 1–12). New York: Springer.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.) (1982). *Judgment under uncertainty: heuristics and biases*. New York: Press Syndicate of the University of Cambridge.

Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction*, *6*(1), 59–98.

Konold, C., & Kazak, S. (2008). Reconnecting data and chance. *Technology Innovations in Statistics Education*, *2*(1). http://escholarship.org/uc/item/38p7c94v.

Konold, C., & Miller, C. D. (2005). *TinkerPlots: dynamic data exploration*. Emeryville: Key Curriculum.

Konold, C., Harradine, A., & Kazak, S. (2007). Understanding distributions by modeling them. *International Journal of Computers for Mathematical Learning*, *12*, 195–216.

Konold, C., Madden, S., Pollatsek, A., Pfannkuch, M., Wild, C., Ziedins, I., Finzer, W., Horton, N. J., & Kazak, S. (2011). Conceptual challenges in coordinating theoretical and data-centered estimates of probability. *Mathematical Thinking and Learning*, *13* (1&2), 68–86.

Lecoutre, M. (1992). Cognitive models and problem spaces in “purely random” situations. *Educational Studies in Mathematics*, *23*, 557–568.

Lee, H., & Lee, J. (2009). Reasoning about probabilistic phenomenon: lessons learned and applied in software design. *Technology Innovations in Statistics Education*, *3*(2). http://escholarship.org/uc/item/1b54h9s9.

Lehrer, R., Kim, M., & Schauble, L. (2007). Supporting the development of conceptions of statistics by engaging students in measuring and modeling variability. *International Journal of Computers for Mathematical Learning*, *12*, 195–216.

Liu, Y., & Thompson, P. (2007). Teachers’ understandings of probability. *Cognition and Instruction*, *25*(2), 113–160.

Nickerson, R. (2004). *Cognition and chance: the psychology of probabilistic reasoning*. Mahwah: Erlbaum.

Pratt, D. (2005). How do teachers foster students’ understanding of probability? In G. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 171–190). New York: Springer.

Shaughnessy, M. (2003). Research on students’ understandings of probability. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 216–226). Reston: National Council of Teachers of Mathematics.

Stohl, H., & Tarr, J. E. (2002). Developing notions of inference using probability simulation tools. *Journal of Mathematical Behavior*, *21*(3), 319–337.

Watson, J. (2005). The probabilistic reasoning of middle school students. In G. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 145–170). New York: Springer.

Commentary on Perspective I: The Humanistic Dimensions of Probability

Bharath Sriraman and Kyeonghwa Lee

The first perspective offered in this book is situated in the historical origins of probability as well as differences in the philosophical bases of probability theory. The four chapters in this section address origins, epistemology, paradoxes, games, modelling of probability and different pedagogical approaches grounded in the *classical*, the *frequentist* or the *subjectivist* approach. The aim of this commentary is not to regurgitate or summarize the content of the chapters, but to ground the first perspective offered in this book within a humanistic framework of mathematics.

The humanist tradition in mathematics education attempts to situate or ground the development of mathematical ideas to the people involved in their conception, and in particular to describe the motivation and context within which the ideas developed. To paraphrase Flusser, the more humanistic aspects of mathematics are the ones that are more fruitful in the sense that "Euler's imperfection (which turns an otherwise perfect piece of workmanship into a work of art), raises many new questions" (p. 5). The origins of probability can be traced to old myths in different cultures in which divination or gambling games are mentioned, often resulting in a catastrophic loss or gain. For example, in the Indian epic Mahabharata (∼1500 B.C.), a kingdom is lost in a dice game which results in exile for the losers. This is not the only instance of "gambling" in ancient Indian culture. The Rig Veda, an old religious scripture contains what is called the Gamblers (Ruin) Hymn in which the dire consequences of playing dice games are extolled in the words, *Cast on the board, like lumps of magic charcoal, though cold themselves they burn the heart to ashes!* Other ancient cultures like the House of Ur in Babylon also developed board games, such as precursors to what is popularly known as backgammon today. Native American cultures in North and South America also independently developed a rich tradition of games

B. Sriraman (✉)
The University of Montana, Missoula, USA
e-mail: SriramanB@mso.umt.edu

K. Lee
Seoul National University, Seoul, Korea

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_6,

of chance. In other words, calculating "odds" seems to have been an integral part of human culture regardless of place and time. The gaming industry today is one of the most profitable sectors of the business world.

If one departs from ancient games of chance to the work coming out of post-Renaissance Europe, the academic strains of what we call probability theory today are found in the works of Luca Pacioli's *Summa de arithmetica geometria proportioni et proportionalia* (1494), as well as Pascal, Fermat, and Jakob Bernoulli's *Ars Conjectandi*" (1713). Borovcnik and Kapadia's opening chapter examines the well-known work of Pascal and Fermat, with reference to de Méré's problem and the Division of Stakes problem. While the former is covered in detail in their chapter, the latter is also worth paying attention to. The Division of Stakes problem if stated in modern terms reads as follows: Suppose a team needs to reach 60 points to win, with each "goal" counting as 10 points, how should the prize money be split up if the score stands at 50–30 and the game cannot be completed? Pacioli's solution would be 5 : 3 and is incorrect but led to the development of a "more accurate" algorithm by Cardano, namely if the score stands at $x : y$ and if "z" is the number needed to win, then the ratio should be $[1 + 2 + 3 + \cdots + (z - y) : 1 + 2 + 3 + \cdots + (z - x)]$. Pacioli's solution would then be corrected to read 6 : 1. However, Cardano was incorrect as well! The reader is invited to devise the correct solution given the tools of modern probability theory at their disposal. Our point here was simply to present the humanistic aspects of developing a solution, which very often consists of trial and errors being corrected over time. The opening chapter also mentions Abraham De Moivre's work on developing "Stirling's formula" but the work that goes unmentioned is the betterment of De Moivre's constant by James Stirling as $\frac{2e}{\sqrt{2\pi}}$, and the historical fluke of De Moivre's work being attributed as Stirling's formula! De Moivre was gracious to let Stirling receive the credit for the betterment of the constant!

The second chapter, also by Borovcnik and Kapadia examines the role of paradoxes in the development of different mathematical concepts. The authors present numerous paradoxes as well as Pacioli's *original* Division of Stakes problem with the historical solutions. De Méré's problem is also revisited along with the problem of the Grand Duke of Tuscany. Numerous other paradoxes are covered in the context of inverse probabilities and the culmination of the chapter is in the unification of many ideas that arose from paradoxes and puzzles that eventually led to the axiomatization of probability theory by Kolmogorov, and the formulation of its central theorems. The paradoxes are also classified as either equal likelihood, expectation, relative frequency, and personal probabilities with the caveat that adopting a particular philosophical stance leads to restrictions in the scope of the applications that are possible. The examples presented are well formulated and serve any interested educator for pedagogical uses.

The last two chapters in this section (Eichler and Vogel; Pfannkuch and Ziedins) examine probability concepts from a mathematical modelling and teaching standpoint. The former present three approaches to modelling situations in probability and question what teachers need to know about probability in order to effectively present and teach the problems. In our reading, a salient feature of the chapter by

Eichler and Vogel is the use of technology like Tinkerplots to simulate modelling situations with data, to eventually lead to the arrival of an "objective" understanding of randomness based on empirical data. The chapter by Pfannkuch and Ziedins extolls the virtues of a data driven approach for teaching probability with three examples to persuade the reader. The classical and frequentist approaches are critiqued as leading to confusion whereas a data-driven modelling approach could circumvent common student misconceptions. To this end, it may be of interest for the statistics education community to realize that modelling and data-driven approaches have been tried in the U.S. in some of the high school reform based curricula sponsored by the National Science Foundation. In the case of Montana, the Systemic Initiative for Montana Mathematics and Science (SIMMS) project resulted in well formulated pilot tested modules containing "real world data". However, this data was perceived as being a little too real (and grim) by some of the public and politicians who used this perception to speak against such approaches to teaching mathematics at high schools. This may serve as the caveat emptor for those that advocate a "radical" approach to the teaching of probability in schools. Sanitized mathematics still fits the world view of most policy makers and too many deviations [pun intended] in the direction of real data may not result in the pedagogical pay out hoped for by these authors!

References

Flusser, P. (1993). A triptych of pythagorean triples. In A. White (Ed.), *Essays in humanistic mathematics*, Washington: Mathematical Association of America.

Perspective II: Psychology

Probabilistic Thinking: Analyses from a Psychological Perspective

Wim Van Dooren

A sophisticated understanding of probability and risk seems evident to become a must in various specialized areas such as medics, economics, logistics, or insurance. But it is equally important in many other domains of modern science such as subatomic physics and in the study of evolution, so that one can easily argue that it is crucial to scientific literacy. Nowadays, however, the ability to interpret and critically evaluate stochastic phenomena that people encounter in diverse contexts is considered essential not only for scientists but for everyone. Being statistically literate (Gal 2002) seems necessary in order to act as an informed and critical citizen in modern society (Rumsey 2002; Utts 2003).

Not only its societal importance turned probabilistic reasoning into a rich domain of research. Probability received interest by researchers from a variety of fields for several other reasons. From a curricular point of view, it is a field having strong links to other areas of mathematics (such as proportional reasoning, combinatorics, multiplicative reasoning, and being closely linked to formal logic as well) while not completely defined by them: Probability is additionally characterized by a focus on situations in which the result is variable and characterized by randomness, uncertainty and outcomes being independent of previous outcomes. Altogether, these characteristics lead to unpredictability, particularly of specific outcomes. So, unlike in other mathematical domains (for instance, I can easily and convincingly show that the area of a circle is not doubled if its diameter is doubled) it is often much more difficult to validate one's assumptions of a probabilistic situation (for instance, the idea that the chance of getting at least one six is doubled if the number of die rolls is doubled).

Besides mathematical challenges, probabilistic situations often also pose emotional challenges. Very often, probabilistic situations are not merely neutral to the problem solver, as the outcomes have a particular personal relevance and emotional, societal or material value: One strategy of playing a game may lead to a larger

W. Van Dooren (✉)
Centre for Instructional Psychology and Technology, University of Leuven, Leuven, Belgium
e-mail: Wim.VanDooren@ppw.kuleuven.be

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_7,

chance of success of winning a valuable prize, the implementation of a certain diagnostic screening may lead to the early detection of a rare disease, but with a risk of showing false positives (and sometimes even more false than true positives).

A final characteristic that makes probability interesting as a domain is that people apparently perform notoriously badly in a wide variety of tasks that require probabilistic thinking to arrive at a normatively correct answer. Systematic errors have been observed in a variety of ages and expertise levels, and also history is full of nice paradoxes and great mathematicians making mistakes (e.g. Székely 1986; Van Dooren et al. 2003).

For all these reasons, it is not very surprising that probability is one of the domains in which psychological research about human reasoning really feels at home. In the current section, chapters will provide several psychological perspectives on the understanding (and/or misunderstanding) of probability.

One of the major lines of debate that run through the various chapters is whether humans can be considered as unbiased, natural-born statisticians, or rather as biased, error-prone probabilistic thinkers. Indeed, probabilistic reasoning seems one of the major battlefields on which the so-called rationality debate takes place. Research originating from the heuristics-and-biases program that originated in the 1970s and 1980s (Kahneman et al. 1982) showed that students fail in solving various simple to rather complex probabilistic problems, due to the strong influence of a limited number of heuristics that may be useful in a variety of situations but also systematically bias human reasoning. These heuristics are often related to dual process theories of reasoning (e.g. Kahneman 2002) and assumed to have a quick, relatively effortless, intuitive nature, as opposed to the slower and more effortful analytic reasoning. The chapters by Chiesi and Primi, and by Ejersbo and Leron explicitly relate to this dual distinction, while also Bennett and Savard build on a distinction between a primitive and a more rational conception. Other chapters by Meder and Gigerenzer, by Martignon, and by Brase, Sherri Martinie and Castillo-Garsow rely on the fact that performance is substantially higher when the same problems are expressed in terms of natural frequencies to argue that humans are essentially good statistical thinkers, a characteristic they evolutionary share with other animals, and which is often called ecological rationality (Todd et al. 2012). Both camps have made appealing claims, but seem to be challenged by the observation that in some circumstances such as situations like the Monty Hall Dilemma, pigeons have been shown to react in an unbiased manner and outperform humans who are frequently biased in this notoriously difficult problem (Herbranson and Schroeder 2010).

Regardless of the radically different points of departure on human rationality in approaching probabilistic situations, the various chapters are reconcilable when it comes to the implications, although this may not seem to be so at first sight, just like optimists and pessimists seem irreconcilable in claiming that the glass of wine is half full or half empty: Both would agree that there is room for more wine in the glass, and even that more wine is desirable. In a similar way, the descriptive approach of probabilistic reasoning may differ substantially, but the prescriptive one—and actually an educational perspective is already taken here—does not: All chapters acknowledge that the existing conceptions in students should be taken as the starting point, and productive bridging approaches should lead students towards normatively correct approaches (and away from what are called preconceptions, alternative conceptions, misconceptions, etc., see the chapter by Savard for a discussion).

So, although taking a psychological perspective in describing probabilistic reasoning, the various chapters do not escape from an educational perspective. This is not surprising, as a thorough understanding of reasoning mechanisms and the origins of prior conceptions may also lead to an engineering of these mechanisms and conceptions. One way in which this happens in several of the chapters (Meder and Gigerenzer, Martignon, Brase, Martini and Castillo-Garsow, Ejersbo and Leron, Abrahamson) is through the use of specifically designed external representations that perceptually highlight probabilities or the mathematical structure underlying complex Bayesian situations.

The fact that people perform worse than expected in probabilistic tasks is not only explained in cognitive terms. Several chapters also refer to psychological mechanisms in the affective field, whose impact on students' responses should be acknowledged as well. Merely the different ideas that students may have of what it means to express a probability, or to denote an event as being random, will have an impact on his or her interpretation of a probabilistic task and thus on the final response. Also, as Chiesi and Primi indicate, it may happen that people possess all required "mindware" (rules and procedures) but perform suboptimally because of superstitious thinking. Bennett also draws attention to the affective domain, by showing how subjective but apparently widely shared preferences (for not acting rather than acting, for instance), distortions in memory (for negative results obtained by action versus inaction, for instance) and emotions of aversion related to regret lead to wrong choices in the Monty Hall Dilemma.

Finally, a sociocultural perspective is brought forward in several of the chapters, implicitly in most, and explicitly by Savard: To what extent can we expect that students use their mathematical insights when approaching probabilistic situations, and to do this consistently? It is very well possible that students are able to apply all required mathematical concepts and procedures to arrive at a normatively correct decision in a game of chance, but at that moment do not consider this to be relevant to the particular situation at that particular moment. In that sense, students' difficulties could be translated into a transfer problem. Lobato's (2003) 'actor oriented transfer' perspective examines the processes by which actors form personal relations of similarity across situations. The actor-oriented transfer perspective argues that the basis for transfer is the learner's often idiosyncratic construal of similarity rather than objectively given similar 'elements' in the environment. Considered in this way, the major educational challenge would be to encourage students to recognize situations in which their probabilistic understanding is applicable, and to enhance their inclination to actually rely on this understanding. So once more, the psychological perspective on probabilistic reasoning and the educational perspective seem to be two sides of the same coin.

References

Gal, I. (2002). Adults' statistical literacy: meaning, components, responsibilities. *International Statistical Review*, *70*, 1–51.

Herbranson, W. T., & Schroeder, J. (2010). Are birds smarter than mathematicians? Pigeons (Columba livia) perform optimally on a version of the Monty Hall Dilemma. *Journal of Comparative Psychology*, *124*(1), 1–13.

Kahneman, D. (2002). Maps of bounded rationality: a perspective on intuitive judgement and choice. Nobel Prize Lecture, December 8 [On-line]. Available: http://nobelprize.org/economics/laureates/2002/kahnemann-lecture.pdf.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgement under uncertainty: heuristics and biases*. Cambridge: Cambridge University Press.

Lobato, J. (2003). How design experiments can inform a rethinking of transfer and vice versa. *Educational Researcher*, *32*(1), 17–20.

Rumsey, D. J. (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education*, *10*(3). Retrieved from www.amstat.org/publications/jse/v10n3/rumsey2.html.

Székely, G. J. (1986). *Paradoxes in probability theory and mathematical statistics*. Dordrecht: Reidel.

Todd, P. M., Gigerenzer, G., & the ABC Research Group (2012). *Ecological rationality: intelligence in the world*. New York: Oxford University Press.

Utts, J. M. (2003). What educated citizens should know about statistics and probability. *American Statistician*, *57*, 74–79.

Van Dooren, W., De Bock, D., Depaepe, F., Janssens, D., & Verschaffel, L. (2003). The illusion of linearity: expanding the evidence towards probabilistic reasoning. *Educational Studies in Mathematics*, *53*, 113–138.

Statistical Thinking: No One Left Behind

Björn Meder and Gerd Gigerenzer

Abstract Is the mind an "intuitive statistician"? Or are humans biased and error-prone when it comes to probabilistic thinking? While researchers in the 1950s and 1960s suggested that people reason approximately in accordance with the laws of probability theory, research conducted in the heuristics-and-biases program during the 1970s and 1980s concluded the opposite. To overcome this striking contradiction, psychologists more recently began to identify and characterize the circumstances under which people—both children and adults—are capable of sound probabilistic thinking. One important insight from this line of research is the power of representation formats. For instance, information presented by means of natural frequencies, numerical or pictorial, fosters the understanding of statistical information and improves probabilistic reasoning, whereas conditional probabilities tend to impede understanding. We review this research and show how its findings have been used to design effective tools and teaching methods for helping people—be it children or adults, laypeople or experts—to reason better with statistical information. For example, using natural frequencies to convey statistical information helps people to perform better in Bayesian reasoning tasks, such as understanding the implications of diagnostic test results, or assessing the potential benefits and harms of medical treatments. Teaching statistical thinking should be an integral part of comprehensive education, to provide children and adults with the risk literacy needed to make better decisions in a changing and uncertain world.

B. Meder (✉) · G. Gigerenzer
Center for Adaptive Behavior and Cognition (ABC), Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany
e-mail: meder@mpib-berlin.mpg.de

G. Gigerenzer
Harding Center for Risk Literacy, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_8,

1 Introduction

Is the mind an "intuitive statistician"? Or are humans biased and error-prone when it comes to probabilistic thinking? Several studies have addressed these questions, giving very different answers. Some research suggests that, by and large, people's intuitive inferences approximately correspond to the laws of probability calculus, mirroring the Enlightenment view that the laws of probability are also the laws of the mind (Daston 1988). Other studies indicate that people have severe difficulties with probability judgments, which has led researchers to conclude that the human mind is not build to reason with probabilities (Tversky and Kahneman 1974).

Much of the work used to corroborate either of the two positions has focused on simple statistical inference problems. Consider a woman applying a home pregnancy test kit. If she tests positive, the probability that she is pregnant increases. Conversely, if the test is negative, the probability that she is pregnant decreases. However, because such tests are not perfectly reliable, these inferences are not certain, but can only be probabilistic. For instance, the sensitivity of home test kits (probability of obtaining a positive test given pregnancy) can be as low as 75 % when applied by inexperienced people (Bastian et al. 1998). This means that out of 100 women who are pregnant, only 75 would test positive, while 25 would test negative. At the same time, such tests are not perfectly specific, meaning that a woman who is not actually pregnant may nevertheless get a positive test result (Bastian et al. 1998).

Situations like this are called *Bayesian problems*, since probability theory, and Bayes' rule in particular, serve as a reference for examining how people revise their beliefs about some state of the world (e.g., being pregnant) in the light of new evidence (e.g., a test result). What are the determinants and limitations of sound reasoning in such tasks? We review research on probabilistic reasoning and show how its findings have been used to design effective tools and teaching methods for helping people—be it children or adults, laypeople or experts—to reason more appropriately with statistical information. Our discussion centers on one of psychologists' *drosophila* for investigating probabilistic thinking, an elementary form of Bayesian reasoning requiring an inference from a single, binary observation to a single, binary hypothesis. Many studies offer a pessimistic view on people's capacity to handle such problems, indicating that both John Q. Public and experts have severe difficulties with them (Kahneman and Tversky 1973; Tversky and Kahneman 1974). However, more recent studies have provided novel insights into the circumstances under which people take the normatively relevant variables into account and are able to solve such problems (Gigerenzer and Hoffrage 1995). Instead of emphasizing human errors, the focus is shifted to human engineering: What can (and need) be done to help people with probabilistic inferences?

One way to foster reasoning with statistical information is to convey information in a transparent and intuitive manner. For instance, a number of studies show that certain frequency formats strongly improve the reasoning of both laypeople in the laboratory (Cosmides and Tooby 1996; Gigerenzer and Hoffrage 1995) and experts outside it (Gigerenzer et al. 2007; Hoffrage and Gigerenzer 1998; Labarge et al.

2003). Drawing on these findings, effective methods have been developed to improve people's ability to reason better with statistical information (Sedlmeier 1999; Sedlmeier and Gigerenzer 2001). These studies show that sound probabilistic thinking is not a mysterious gift out of reach for ordinary people, but that one can learn to make better inferences by using the power of representation: "Solving a problem simply means representing it so as to make the solution transparent" (Simon 1969, p. 153).

2 Bayesian Reasoning as a Test Case of Probabilistic Thinking

In situations like the pregnancy test, a piece of evidence is used to revise one's opinion about a hypothesis. For instance, how does a positive test change the probability of being pregnant? From a statistical point of view, answering this question requires taking into account the prior probability of the hypothesis H (i.e., the probability of being pregnant *before* the test is applied) and the likelihood of obtaining datum D under each of the hypotheses (i.e., the likelihood of a positive test when pregnant and the likelihood of a positive test when not pregnant). From this, the posterior probability of being pregnant given a positive test, $P(H|D)$, can be computed by Bayes' rule:

$$P(H|D) = \frac{P(D|H) \times P(H)}{P(D|H) \times P(H) + P(D|\neg H) \times P(\neg H)} = \frac{P(D|H) \times P(H)}{P(D)} \quad (1)$$

where $P(H)$ denotes the hypothesis' prior probability, $P(D|H)$ denotes the probability of observing datum D given the hypothesis is true, and $P(D|\neg H)$ denotes the probability of D given the hypothesis is false.

2.1 Is the Mind an "Intuitive Statistician"?

Bayes' rule is an uncontroversial consequence of the axioms of probability theory. Its status as a descriptive model of human thinking, however, is not at all self-evident. In the 1950s and 1960s, several studies were conducted to examine to what extent people's intuitive belief revision corresponds to Bayes' rule (e.g., Edwards 1968; Peterson and Miller 1965; Peterson et al. 1965; Phillips and Edwards 1966; Rouanet 1961). For instance, Phillips and Edwards (1966) presented subjects with a sequence of draws from a bag containing blue and red chips. The chip was drawn from either a predominantly red bag (e.g., 70 % red chips and 30 % blue chips) or a predominantly blue bag (e.g., 30 % red chips and 70 % blue chips). Given a subject's prior belief about whether it came from a predominantly red or predominantly blue bag, the question of interest was whether subjects would update their beliefs in accordance with Bayes' rule.

This and several other studies showed that participants took the observed evidence into account to some extent, but not as extensively as prescribed by Bayes' rule. “It turns out that opinion change is very orderly and usually proportional to numbers calculated from Bayes’s theorem—but it is insufficient in amount.” (Edwards 1968, p. 18). Edwards and colleagues called this observation *conservatism*, meaning that people shifted their probability estimates in the right direction, but did not utilize the data as much as an idealized Bayesian observer would. While there was little disagreement on the robustness of the phenomenon, different explanations were put forward (see Edwards 1968, for a detailed discussion). One idea was that conservatism results from a *misaggregation* of information, such as a distorted integration of priors and likelihoods. Other researchers suggested that the inference process itself principally follows Bayes’ rule, but a *misperception* of the data-generating processes and the diagnostic value of data would result in estimates that are too conservative. Another idea was that the predominantly used book, bag, and poker chip tasks were too artificial to draw more general conclusions and that people outside the laboratory would have less difficulty with such inferences.

Despite some systematic discrepancies between Bayes’ rule and people’s inferences, the human mind was considered to be an (albeit imperfect) “intuitive statistician”: “Experiments that have compared human inferences with those of statistical man show that the normative model provides a good first approximation for a psychological theory of inference.” (Peterson and Beach 1967, p. 43).

2.2 Is the Human Mind Biased and Error-Prone when It Comes to Probabilistic Thinking?

Only a few years later, other researchers arrived at a very different conclusion: “In making predictions and judgments under uncertainty, people do not appear to follow the calculus of chance or the statistical theory of prediction.” (Kahneman and Tversky 1973, p. 237).

What had happened? Other studies had been conducted, also using probability theory as a normative (and potentially descriptive) framework. However, people’s behavior in these studies seemed to be error-prone and systematically biased (Bar-Hillel 1980; Tversky and Kahneman 1974; Kahneman and Tversky 1972, 1973; Kahneman et al. 1982).

One (in)famous example of such an inference task is the so-called “mammography problem” (adapted from Eddy 1982; see Gigerenzer and Hoffrage 1995):

> The probability of breast cancer is 1 % for a woman at the age of 40 who participates in routine screening. If a woman has breast cancer, the probability is 80 % that she will get a positive mammography. If a woman does not have breast cancer, the probability is 9.6 % that she will also get a positive mammography. A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer? ___%

Here, the hypothesis in question is whether the woman has breast cancer given the base rate of cancer (1 %) and the likelihood of obtaining a positive test result for women with cancer (80 %) or without cancer (9.6 %).

From the perspective of statistical inference, the problem is simple. The relevant pieces of information are the prior probability of disease, $P(\text{cancer})$, the probability of obtaining a positive test result for a woman having cancer, $P(T+|\text{cancer})$, and the probability of obtaining a positive test result for a woman having *no* cancer, $P(T+|\text{no cancer})$. From this information, the probability of breast cancer given a positive test result, $P(\text{cancer}|T+)$, can be easily computed by Bayes' rule:

$$\begin{aligned} P(\text{cancer}|T+) &= \frac{P(T+|\text{cancer}) \times P(\text{cancer})}{P(T+)} \\ &= \frac{P(T+|\text{cancer}) \times P(\text{cancer})}{P(T+|\text{cancer}) \times P(\text{cancer}) + P(T+|\text{no cancer}) \times P(\text{no cancer})} \\ &= \frac{0.8 \times 0.01}{0.8 \times 0.01 + 0.096 \times 0.99} = 0.078 \approx 8\ \%. \end{aligned}$$

Thus, the probability that a woman with a positive mammogram has cancer is about 8 %. In stark contrast, empirical research shows that both health care providers and laypeople tend to give much higher estimates, often around 70 %–80 % (Casscells et al. 1978; Eddy 1982; Gigerenzer and Hoffrage 1995; Hammerton 1973; for reviews, see Bar-Hillel 1980; Koehler 1996a). These overestimates were interpreted as evidence for *base rate neglect*, suggesting that people did not seem to take base rate information (e.g., prior probability of cancer) into account to the extent prescribed by Bayes' rule.

These and similar findings were bad news for the notion of "statistical man" (Peterson and Beach 1967). After all, from a formal point of view, this problem is essentially as easy as it can get: A binary hypothesis space (cancer vs. no cancer), a single datum (positive test result), and all the information necessary to make the requested inference is available. If people failed to solve such an apparently simple problem, it seemed clear that humans lack the capacity to reason in accordance with Bayes' rule: "In his evaluation of evidence, man is apparently not a conservative Bayesian: he is not Bayesian at all." (Kahneman and Tversky 1972, p. 450).

3 The Power of Presentation Formats

While researchers in the 1950s and 1960s believed that people reason approximately in accordance with the laws of probability theory, albeit conservatively, studies conducted in the heuristics-and-biases program during the 1970s and 1980s concluded the opposite. To overcome this conceptual impasse, psychologists more recently began to identify and characterize the circumstances under which people—both children and adults—are capable of sound probabilistic thinking.

Gigerenzer and Hoffrage (1995) argued that one crucial component is the link between cognitive processes and information formats. For instance, mathematical operations like multiplication and division are hard with numbers represented as Roman numerals, but comparatively easy with Arabic numerals. The theoretical general point is that the representation does part of the job and can facilitate reasoning.

Gigerenzer and Hoffrage (1995; see also Cosmides and Tooby 1996) compared the above version of the mammography problem, in which numerical information is presented in terms of conditional probabilities, with a version in which the same information is expressed in terms of *natural frequencies*:

> 10 out of every 1,000 women at the age of 40 who participate in routine screening have breast cancer. 8 of every 10 women with breast cancer will get a positive mammography. 95 out of every 990 women without breast cancer will also get a positive mammography. Here is a new representative sample of women at the age of 40 who got a positive mammography in routine screening. How many of these women do you expect to actually have breast cancer? ___ out of ___ .

When the problem was presented this way, a sharp increase in correct ("Bayesian") answers was obtained. For instance, in the mammography problem, 16 % of participants gave the correct solution when information was presented as probabilities, as opposed to 46 % in the natural frequency version. Similar results were obtained across 15 different problems (Gigerenzer and Hoffrage 1995).

Why is this? One general idea is that the human mind is better adapted to reason with frequency information, as this is the "raw data" we experience in our daily life and have adapted to throughout evolutionary history (Cosmides and Tooby 1996). For instance, over the course of professional life, a physician may have diagnosed hundreds of patients. The doctor may have experienced that out of 1,000 patients, few will actually have a certain disease (e.g., 10 out of 1,000) and that most of those (e.g., 8 out of 10) had a certain symptom. But the physician may also have experienced that of the many people who do not have the disease some also show the symptom (e.g., 95 out of the 990 without the disease). Now, if confronted with a new patient showing the symptom, the physician will know from past experience that only 8 out of 103 patients with the symptom actually have the disease. Thus, while aware of the fact that most people with the disease show the symptom, the physician is also well aware of the relatively high number of false positives (i.e., people without the disease but having the symptom) resulting from the low base rate of the disease (Christensen-Szalanski and Bushyhead 1981). If the relation between diagnostic cues and base rates is experienced this way, people's Bayesian reasoning has been shown to improve (Christensen-Szalanski and Beach 1982; see also Koehler 1996a; Meder and Nelson 2012; Meder et al. 2009; Nelson et al. 2010).

This leads directly to a second, more specific explanation: Natural frequencies facilitate probabilistic inference because they simplify the necessary calculations. A different way of computing a posterior probability according to Bayes' rule (Eq. (1)) is

$$P(D|H) = \frac{N(H \cap D)}{N(D)} \tag{2}$$

where $N(H \cap D)$ denotes the number of instances in which the hypothesis is true and the datum was observed (e.g., patients with disease and symptom) and $N(D)$ denotes the total number of cases in which D was observed (e.g., all patients with the symptom). (The equivalence of Eqs. (1) and (2) follows from the axioms of probability theory, according to which $P(H \cap D) = P(D|H)P(H)$.) For the mammography problem, Eq. (2) yields

$$P(\text{cancer}|\text{positive test}) = \frac{N(\text{cancer} \cap \text{positive test})}{N(\text{positive test})} = \frac{8}{8+95} = 0.078 \approx 8\ \%$$

where N(cancer ∩ positive test) denotes the number of women with cancer and a positive mammogram, and N(positive test) denotes the total number of cases with a positive test result.

Although this notation is mathematically equivalent to Eq. (1), it is not psychologically equivalent. Representing the probabilistic structure of the environment in terms of natural frequencies simplifies Bayesian reasoning because the required mathematical operation can be performed on natural numbers rather than on normalized fractions (i.e., probabilities). And because the base rate information is already contained in the natural frequencies that enter the computation, it must not be explicitly considered in the calculations. Thus, the representation does part of the job.

3.1 What Are Natural Frequencies and why Do They Simplify Bayesian Reasoning?

It is important to understand what natural frequencies are and what they are not. For instance, the findings of Gigerenzer and Hoffrage (1995) have sometimes been interpreted as suggesting that *any* type of frequency information facilitates probabilistic reasoning (e.g., Evans et al. 2000; Girotto and Gonzalez 2001; Johnson-Laird et al. 1999; Lewis and Keren 1999). However, the claim is that natural frequencies improve Bayesian reasoning (Gigerenzer and Hoffrage 1999; Hoffrage et al. 2002). By contrast, relative (normalized) frequencies should not—and did not—facilitate reasoning (Gigerenzer and Hoffrage 1995).

To understand why natural frequencies—but not frequencies in general—facilitate Bayesian reasoning, one needs to distinguish between *natural sampling* and *systematic sampling* (Kleiter 1994). The important difference is that natural samples preserve base rate information, whereas systematic sampling—the usual approach in scientific research—fixes base rates a priori. For example, if we take a random sample of 1,000 women from the general population, we will find that 10 of these 1,000 women have breast cancer (Fig. 1, left). Furthermore, we would observe that out of the 10 women with cancer, 8 will get a positive mammogram, and that 95 out of the 990 women without cancer will get a positive mammogram. The probability of women having cancer given a positive test result can now be easily calculated by comparing the (natural) frequency of women with cancer *and* a positive test result to the overall number of women with a positive mammogram (Fig. 1, left), namely, $8/(95 + 8)$ (Eq. (2)). As mentioned earlier, because these numbers already contain base rate information, it is unnecessary to explicitly consider this information when evaluating the implications of a positive test result.

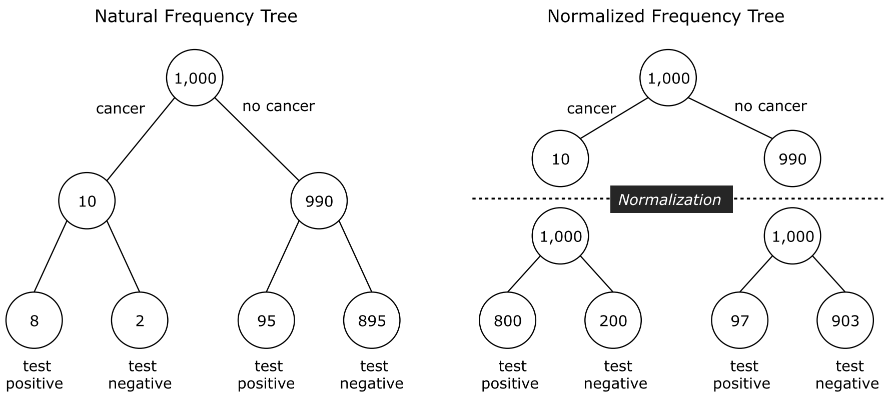


Fig. 1 A natural frequency tree (*left*) and a normalized frequency tree (*right*). The four numbers at the *bottom of the left tree* are natural frequencies; the four numbers at the *bottom of the right tree* are not. The natural frequency tree results from natural sampling, which preserves base rate information (number of women with cancer vs. without cancer in the population). In the normalized tree, the base rate of cancer in the population (10 out of 1,000) is normalized to an equal number of people with and without cancer (e.g., 1,000 women with and 1,000 women without cancer). In the natural frequency tree, the posterior probability of $P(\text{cancer} \mid \text{test positive})$ can be read from the tree by comparing the number of people with cancer and a positive test with the total number of positive test results ($8/(8+95)$). This does not work when base rates have been fixed a priori through systematic sampling (*right tree*)

However, this calculation is only valid with natural frequencies. It does not work with normalized (relative) frequencies resulting from systematic sampling, in which the base rate of an event in the population is normalized. For instance, to determine the specificity and sensitivity of a medical test, one might set up an experiment with an equal number of women with and without cancer (e.g., 1,000 women in each group) and apply the test to each woman (Fig. 1, right). The results make it possible to assess the number of true positives (women with cancer who get a positive test result) and false positives (women without cancer who get a positive test result) as well as the number of true negatives (women without cancer who get a negative test result) and false negatives (women with cancer who get a negative test result). This method is appropriate when evaluating test characteristics (e.g., sensitivity and specificity of a medical test) because it ensures that samples of women with and without cancer are large enough for statistically reliable conclusions to be drawn. However, when used to make diagnostic inferences, base rates must be explicitly reintroduced into the calculation (via Eq. (1)).

4 Can Children Solve Bayesian Problems?

The classic view endorsed by Piaget and Inhelder (1975) holds that young children lack the specific capacities necessary for sound probability judgments. For instance, the application of a combinatorial system and the capacity to calculate proportions are prerequisites for reasoning with probabilities, abilities that are assumed not to develop before the age of 11 or 12.

On the other hand, studies indicate that even young children have basic intuitions regarding probability concepts (Fischbein et al. 1970; Girotto and Gonzalez 2008; Yost et al. 1962). Brainerd (1981) demonstrated some basic skills in children for understanding the outcomes of sampling processes given a certain distribution of events in a reference class. In one experiment, 10 tokens of two different colors were placed in an opaque container (e.g., 7 red and 3 black tokens). When children were asked to predict the outcome (e.g., black or red token) of a series of random draws, their predictions for the first draw usually preserved the ordering of the sampling probabilities. Although children's predictions were not consistent with the current sampling proportions in subsequent trials, when explicitly probed about the relative frequencies before each prediction, the performance of both younger children (preschoolers age 4 to 5) and older children (2nd and 3rd graders) improved (Brainerd 1981, Studies 6 and 12). The observed developmental trajectories also point to the link between probability judgments and basic cognitive processes, such as the storage, retrieval, and processing of frequency information in and from memory. A similar development across age groups was observed by Girotto and Gonzalez (2008), who showed that starting around the age of 5, children are sensitive to new evidence when betting on random events (i.e., draws from a bag containing chips) and can guess what outcome is more likely.

Using natural frequencies, can children also make more quantitative inferences in Bayesian reasoning tasks? Zhu and Gigerenzer (2006; see also Multmeier 2012) investigated children's (4th, 5th, and 6th graders) capacity to reason in accordance with Bayes' rule by using similar tasks to the mammography problem. One of these, for instance, was the "red nose" problem. This is the problem stated in terms of conditional probabilities (expressed as percentages):

> Pingping goes to a small village to ask for directions. In this village, the probability that the person he meets will lie is 10 %. If a person lies, the probability that he/she has a red nose is 80 %. If a person doesn't lie, the probability that he/she also has a red nose is 10 %. Imagine that Pingping meets someone in the village with a red nose. What is the probability that the person will lie?

The solution to this question can be inferred using Bayes' rule (Eq. (1)), which gives a posterior probability of 47 %.

Are children able to solve such a problem? The answer is no: None of the 4th, 5th, or 6th graders could solve such a problem when information was presented in terms of conditional probabilities. Of the adult control sample, about half of the participants were able to solve the problem (Fig. 2).

The main question was whether children would do better if the same information was presented in terms of natural frequencies. This is the same red nose problem in a natural frequency format:

> Pingping goes to a small village to ask for directions. In this village, 10 out of every 100 people will lie. Of the 10 people who lie, 8 have a red nose. Of the remaining 90 people who don't lie, 9 also have a red nose. Imagine that Pingping meets a group of people in the village with red noses. How many of these people will lie? ____ out of ____.

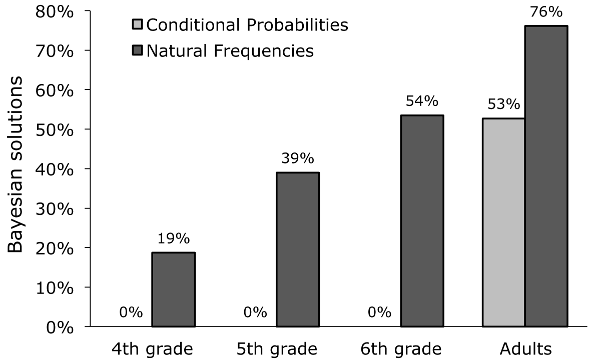


Fig. 2 Percentage of correct solutions in Zhu and Gigerenzer (2006; results aggregated across Studies 1 and 2). No child was able to solve a Bayesian reasoning task when information was presented probabilities. Performance was much better when reasoning with natural frequencies

The findings show that natural frequencies can help children to understand and solve problems that are otherwise beyond their skills (Fig. 2). The results also show a strong trend with age, similar to results of other studies asking children for categorical probability judgments (Brainerd 1981; Girotto and Gonzalez 2008).

5 Bayesian Reasoning Outside the Laboratory

Apart from conducting laboratory studies on Bayesian reasoning (in which the participants are often students), researchers have also examined experts' capacity to reason in accordance with Bayes' rule. This is important, given that making inferences from statistical data is an important part of decision making in many areas, such as medicine and the law.

However, research shows that statistical illiteracy is a widespread phenomenon among experts as well (Gigerenzer et al. 2007; Gigerenzer and Gray 2011). For instance, Wegwarth et al. (2012) found in a survey in the United States that most primary care physicians have severe difficulties understanding which statistics are relevant to assessing whether screening saves lives. When examining how general practitioners interpret the results of a diagnostic test (probability of endometrial cancer given a transvaginal ultrasound), Steurer et al. (2002) found that most physicians strongly overestimated the probability of disease given the positive test result.

Another example is probabilistic thinking in the law. Research indicates that—similar to health care providers—jurors, judges, and lawyers often confuse and misinterpret statistical information (Gigerenzer 2002; Kaye and Koehler 1991; Koehler 1996b; Thompson and Schumann 1987). Overall, research on reasoning with probabilities in legal contexts mirrors the difficulties observed in research on probability judgments in general, including a tendency to neglect base rate information but also the opposite, giving too little weight to the evidence (Gigerenzer 2002; Thompson and Schumann 1987).

5.1 Improving Physicians' Diagnostic Reasoning Through Natural Frequencies

Hoffrage and Gigerenzer (1998) investigated the impact of presentation format in the diagnostic reasoning of experienced physicians with different backgrounds (e.g., gynecologists, radiologists, and internists). Physicians were presented with diagnostic inference tasks, such as estimating the probability of breast cancer given a positive mammogram or the probability of colorectal cancer given a positive Hemoccult test. For each problem, the relevant information on base rates, the probability of a positive test result given the disease, and the probability of a positive test given no disease was provided either in terms of probabilities or as natural frequencies. The results showed a dramatic difference between the two presentation formats, increasing from an average of 10 % correct solutions when reasoning with probabilities to an average of 46 % correct solutions when reasoning with natural frequencies. These findings show that experts, too, find it much easier to make probabilistic diagnostic inferences when information is presented in terms of natural frequencies than with probabilities.

In Labarge et al.'s (2003) study with 279 members of the National Academy of Neuropsychology, the physicians' task was to estimate the probability that a patient has dementia, given a positive dementia screening score. Information on the base rate of the disease and the test characteristics were provided in terms of either probabilities or natural frequencies. When information was conveyed through probabilities, only about 9 % of the physicians correctly estimated the posterior probability. By contrast, if the information was provided in a natural frequency format, 63 % correct answers were obtained.

Bramwell et al. (2006) examined the influence of presentation formats on the interpretation of a screening test for Down syndrome. Their findings show that, with probabilities, only 5 % of obstetricians drew correct conclusions, but, when information was presented as natural frequencies, 65 % gave the correct answer. However, their findings also show that other stakeholders in screening (e.g., pregnant women, midwives) had difficulties drawing correct conclusions, even with frequency information. This points to the importance of systematically training health care providers in reasoning with probabilistic information, possibly by using even more intuitive presentation formats such as visual representations of frequency information (see below).

5.2 Probabilistic Thinking in the Law

Advances in forensic science have made the use of DNA analyses a common practice in legal cases, requiring jurors and judges to makes sense of statistical information presented by the prosecution or the defense. In the O. J. Simpson case, for instance, numerous pieces of genetic evidence from the crime scene were introduced during testimony, such as blood matches between samples from the crime scene and the defendant. Typically, each piece of supposed evidence was presented along with numerical information, such as the probability that the DNA profile of a randomly selected person would match a genetic trace found at the crime scene (so-called "random match probability", see Weir 2007). In one afternoon alone, the prosecution presented the jury with 76 different quantitative estimates (Koehler 1996b).

As in the medical domain, one goal should be to present statistical information in a transparent manner in order to avoid confusion and assist decision makers in making better inferences. This is of particular importance in legal cases, where the prosecutor and the defense may present information in a strategic way to influence the judge or the jury in one way or the other (Gigerenzer 2002; Thompson and Schumann 1987). Lindsey et al. (2003; see also Hoffrage et al. 2000) examined the effects of presenting information on forensic evidence in different presentation formats with a sample of advanced law students and professional jurists. The goal was to compare probabilistic reasoning when numerical information is conveyed through probabilities and when conveyed through natural frequencies. Consider the following example (Lindsey et al. 2003):

> In a country the size of Germany, there are as many as 10 million men who fit the description of the perpetrator. The probability of a randomly selected person having a DNA profile that matches the trace recovered from the crime scene is 0.0001 %. If someone has this DNA profile, it is practically certain that this kind of DNA analysis would show a match. The probability that someone who does not have this DNA profile would match in this type of DNA analysis is 0.001 %. In this case, the DNA profile of the sample from the defendant matches the DNA profile of the trace recovered from the crime scene.

Given this information, the probability that someone has a particular DNA profile if a match was obtained with the trace from the crime scene (i.e., $P(\text{profile} \mid \text{match})$) can be computed according to Bayes' rule (Eq. (1)), yielding a posterior probability of about 9 %. However, given this method of presenting statistical information, only 1 % of law students and 11 % of the jurists were able to derive the correct answer.

Do natural frequencies help understand the numerical information? The same information as above, expressed in terms of natural frequencies, reads as follows:

> In a country the size of Germany, there are as many as 10 million men who fit the description of the perpetrator. Approximately 10 of these men would have a DNA profile that matches the trace recovered from the crime scene. If someone has this DNA profile, it is practically certain that this kind of DNA analysis would show a match. Out of 9,999,990 people who do not have this DNA profile, approximately 100 would be shown to match in this type of DNA analysis. In this case, the DNA profile of the sample from the defendant matches the DNA profile of the trace recovered from the crime scene.

When the statistical information was conveyed this way, significantly more correct answers were obtained (40 % from the law students and 74 % from the professional jurists). This finding is consistent with research on the medical diagnosis task, showing that both legal laypeople (or advanced students, in this case) and experts such as professional jurists can benefit from the use of natural frequencies.

The probative evidence of forensic tests is, of course, not the only factor that plays an important role in legal cases (Koehler 2006). However, presenting evidence so that judges and jurors understand its meaning—and the uncertainties associated with such analyses—is an important prerequisite for making informed decisions.

5.3 Risk Communication: Pictorial Representations

One of the most successful applications of transparent and intuitive presentation formats is risk communication in the health domain. Informed medical decisions—such as whether to participate in a cancer screening program or choosing between alternative treatments—require that both health professionals and patients understand the relevant probabilities.

One example for the importance of understanding quantitative health information concerns the benefits of cancer screening programs, for instance, PSA screening for prostate cancer (Arkes and Gaissmaier 2012). Research shows that both the general public and physicians overestimate the benefits of such programs. For instance, Wegwarth and colleagues (2012) showed that most physicians strongly overestimate the benefits of PSA screening and are led astray by irrelevant statistics. According to a survey with a representative sample of more than 10,000 people from nine countries on the perceived benefits of breast and prostate cancer screening (Gigerenzer et al. 2009), people largely overestimate the benefits of these screening programs. These findings are in stark contrast to the recommendations of health organizations like the U.S. Preventive Services Task Force, who for instance explicitly recommends against PSA-based screening for prostate cancer (Moyer 2012).

What can be done to improve the understanding of the risks and benefits of medical treatments? Using frequencies in either numerical or pictorial form can help people make better, more informed decisions (Akl et al. 2011; Ancker et al. 2006; Edwards et al. 2002; Fagerlin et al. 2005; Gigerenzer and Gray 2011; Gigerenzer et al. 2007; Kurz-Milcke et al. 2008; Lipkus and Hollands 1999). Visualizing statistical information may be especially helpful for people having difficulties in understanding and reasoning with numerical information (Lipkus et al. 2001; Schwartz et al. 1997). Figure 3 gives an example of an *icon array* illustrating the effect of aspirin on the risk of having a stroke or heart attack (Galesic et al. 2009). This iconic representation of simple frequency information visualizes that 8 out of 100 people who do not take aspirin have a heart attack or stroke, as opposed to 7 out of 100 people who do take aspirin. This reduction corresponds to a *relative risk reduction* of about 13 % $[(8-7)/8]$. Although communicating the benefits of treatments through relative risk reductions is common practice—in medical journals as well as in patient counseling and the public media—it has been shown to lead to overestimations of treatment benefits (Akl et al. 2011; Bodemer et al. 2012; Covey 2007; Edwards et al. 2001). Galesic and colleagues (2009) showed that iconic representations help both younger and older adults gain a better understanding of relative risk reductions. Icon arrays were particularly helpful for participants with low numeracy skills (Lipkus et al. 2001; Peters 2008; Schwartz et al. 1997).

Galesic and colleagues used icon arrays to visualize simple frequency information (event rates in treatment and control group) and thus aid people in understanding the meaning of relative risk reduction. Another example of an icon array visualizing information on the benefits and harms of PSA screening is shown in Fig. 4 (cf. Arkes and Gaissmaier 2012). It contrasts two groups of people, men who participate in PSA screening versus men who do not. The icon array provides information on

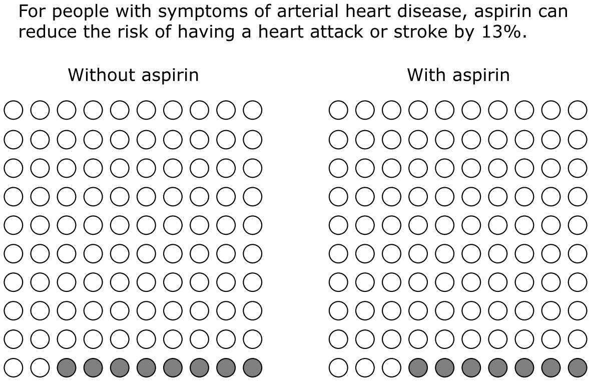


Fig. 3 Icon array used by Galesic et al. (2009) to visualize the effect of taking aspirin on the risk of having a heart attack. The data entail a relative risk reduction of 13 % (from 8 out of 100 to 7 out of 100)

how many men out of 1,000 aged 50 and older are expected to die from prostate cancer in each of the groups, as well as the number of false alarms and number of men diagnosed and treated for prostate cancer unnecessarily. This visual display provides a transparent and intuitive way of communicating the potential benefits and harms of a medical treatment.

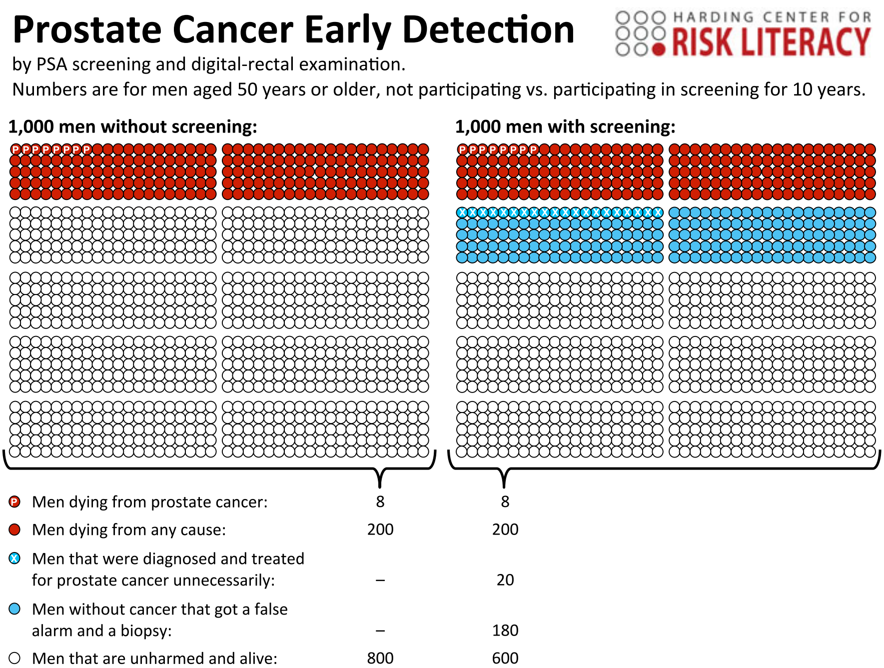


Fig. 4 Example of an icon array visualizing the benefits and harms of prostate-specific antigen (PSA) screening for men age 50 and older. The epidemiological data visualized here is taken from Djulbegovic et al. (2010). Copyright 2012 by the Harding Center for Risk Literacy (www.harding-center.com)

However, one should also note that not all graphical aids are equally effective (Ancker et al. 2006). For instance, Brase (2009) compared icon arrays with two types of Venn diagrams, one with and one without individuating information (i.e., dots in the Venn diagram, with each dot representing an individual person). His findings show a consistent advantage of iconic representations over both types of Venn diagrams, pointing to a special role of iconic representations and the importance of choosing the right visual representation.

6 Teaching Representations

How can the findings from cognitive psychology be used to effectively teach probabilistic thinking? We believe that there are two important lessons to be learned from psychological research. First, when information is numerically presented in terms of probabilities, people have great difficulties in making sound inferences. Second, this difficulty can be overcome by conveying information through natural frequencies rather than (conditional) probabilities. Creating an alternative representation of the problem facilitates much better understanding of—and reasoning with—statistical information.

However, whereas natural frequencies strongly improve people's performance, research also indicates that performance is not perfect. For instance, in the Gigerenzer and Hoffrage (1995) study, natural frequencies elicited around 50 % correct responses. This finding is impressive compared to the number of correct responses when reasoning with conditional probabilities, but performance can be further improved by systematically teaching people how to use the power of presentation formats.

Sedlmeier and Gigerenzer (2001; see also Cole and Davidson 1989; Kurz-Milcke and Martignon 2006; Kurzenhäuser and Hoffrage 2002; Sedlmeier 1999) developed a computer-based tutorial program for teaching Bayesian reasoning. The key feature was to teach participants to translate statistical problems into presentation formats that were more readily comprehensible. The study compared rule-based training (i.e., learning how to plug in probabilities into Bayes' rule) with two types of frequency formats: a natural frequency tree (as in Fig. 1 left) and an icon array (Sedlmeier and Gigerenzer used the term frequency grid). A control group without training served as a base line condition. Both short- and long-term training effects were assessed, as was the generalization to new (elemental Bayesian reasoning) problems.

The basic setup of the training procedure was as follows: First, participants were presented with two text problems (the mammography problem and the sepsis problem; see Fig. 5). In the *rule condition*, participants learned how to insert the probabilities stated in the problem in Bayes' rule (Eq. (1)). In the two frequency representation conditions, participants learned how to translate the given probabilities into a natural frequency tree (Fig. 1, left) or an icon array (Fig. 5).

After being guided through the first two problems, participants were presented with eight further problems. For each problem, their task was to solve the problem

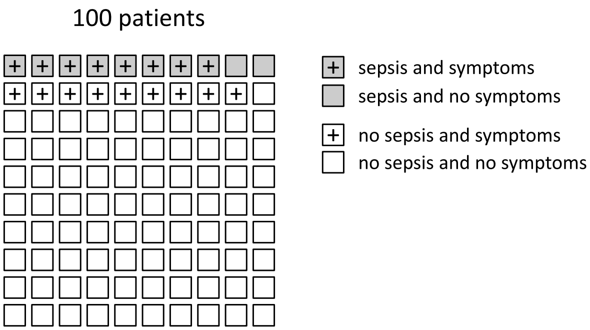


Fig. 5 Icon array similar to the one Sedlmeier and Gigerenzer (2001) used to train participants on translating probabilities into natural frequencies. The task is to infer the posterior probability of sepsis given the presence of certain symptoms, $P(\text{sepsis} \mid \text{symptoms})$. Each square represents one individual. The icon array illustrates the base rate of sepsis in a sample of 100 patients (grey squares) and the likelihood of the symptoms (denoted by a "+") in patients with and without sepsis. The frequencies correspond to the following probabilities: $P(\text{sepsis}) = 0.1$, $P(\text{symptoms} \mid \text{sepsis}) = 0.8$, $P(\text{symptoms} \mid \text{no sepsis}) = 0.1$

(i.e., to compute the posterior probability) by inserting the probabilities into Bayes' rule or by creating a frequency representation (tree vs. icon array, respectively). When participants had difficulties solving these problems, the computer provided help or feedback.

The effectiveness of the training methods was assessed through three post-training sessions (immediate, 1-week follow-up, 5-weeks follow-up). The results showed that teaching participants to translate probabilities into natural frequencies was the most effective method. For instance, prior to training, the median percentage of correct solutions was 0 % (rule condition) and 10 % (natural frequency conditions). When tested immediately after training, the median number of correct solutions was 60 % in the rule condition, 75 % in the icon array condition, and 90 % in the natural frequency tree condition. The most important findings concern the stability of the training effects over time. In the rule-based training, participants' performance was observed to decrease over time; after 5 weeks, the median percentage of correct solutions was reduced to 20 %. For participants who had been trained to use natural frequency formats, by contrast, no such decrease was observed.

7 Conclusions

Thinking and reasoning with statistical information is a challenge for many of us. Many researchers are (or have been) of the opinion that people are severely limited in their capacity of sound probabilistic thinking and fall prey to "cognitive illusions" (Edwards and von Winterfeldt 1986). However, whereas visual illusions may be an unavoidable by-product of the perception system, cognitive illusions are not hard-wired (for a detailed discussion, see Gigerenzer 1996; Kahneman and Tversky 1996). Recent research has revealed insights on how to help people—adults and children, laypeople and experts—to reason better with probabilistic information about risk and uncertainties. The insights gained from this line of research have also been used to inform applied research, such as how to effectively communicate risks and benefits of medical treatments to health professionals and patients. The most important lesson learned from this research is that information formats matter—in fact, they matter strongly.

7.1 Implications for Mathematics Education

Children, even at a young age, have basic intuitions about probability concepts, such as how the proportion of events relates to the outcomes of simple sampling processes (Brainerd 1981; Fischbein et al. 1970; Girotto and Gonzalez 2008; Yost et al. 1962). One goal of mathematics education should be to foster these intuitions from elementary school on, so that they can later serve as a basis for the acquisition of probability calculus.

A promising road to advance such intuitions is to use playful activities, for example, assembling the so-called "tinker cubes" to illustrate and compare different proportions (Martignon and Krauss 2007, 2009; Kurz-Milcke et al. 2008; see also Martigon 2013). Such intuitive representations can also be used to illustrate the relation between feature distributions and simple sampling processes (Kurz-Milcke and Martignon 2006). Later, when children are taught the basics of probability theory, these intuitions can help pupils develop an understanding of probabilities and simple statistical inferences. Teaching concepts like conditional probabilities, in turn, should use real-world examples and capitalize on the power of natural frequencies and visual aids such as icon arrays and trees. Children and adults alike should be taught representations and not merely the application of rules.

7.2 The Art of Decision Making

Making good decisions requires more than just number crunching. First, it is very important to develop an understanding of the very concepts to which the numbers refer. For instance, research on the perceived benefits of PSA screening shows that physicians are often led astray by statistical evidence that is irrelevant to the question of whether someone should participate in screening (Wegwarth et al. 2012, 2011). Here and in other situations, a qualitative understanding of the concepts behind the numerical information is a prerequisite for informed decisions. Second, in many situations, reliable probability estimates are not available and cannot be relied upon in the decision making process. Such situations are called "decision making under uncertainty" as opposed to "decision making under risk" (Knight 1921/2006) and require more (and other) cognitive tools than probability. When making decisions under uncertainty, simple heuristics are the tools that people use—and should use (Gigerenzer et al. 1999, 2011).

We therefore believe that teaching children the basics of decision making should be an integral part of education. For instance, it is important to understand that there is no unique, always-optimal way of making decisions. Some situations may require a deliberative reasoning process based on statistical evidence, while other situations might require relying on gut feelings (Gigerenzer 2007). One step toward learning the art of decision making is to understand that a toolbox of decision-making strategies exists that can help people deal with a fundamentally uncertain world (Gigerenzer and Gaissmaier 2011; Gigerenzer et al. 2011; Hertwig et al. 2012).

7.3 Final Thoughts

Two centuries ago, few could imagine a society where everyone could read and write. Today, although the ability to reckon with risk is as important as reading and writing, few can imagine that the widespread statistical illiteracy can be vanquished. We believe that it can be done and that it needs to be done. Learning how to understand statistical information and how to use it for sound probabilistic inferences must be an integral part of comprehensive education, to provide both children and adults with the risk literacy needed to make better decisions in a changing and uncertain world.

Acknowledgements BM was supported by Grant ME 3717/2 from the Deutsche Forschungsgemeinschaft (DFG) as part of the priority program "New Frameworks of Rationality" (SPP 1516). We thank Christel Fraser and Rona Unrau for editing the manuscript.

References

Akl, E. A., Oxman, A. D., Herrin, J., Vist, G. E., Terrenato, I., Sperati, F., Costiniuk, C., Blank, D., & Schünemann, H. (2011). Using alternative statistical formats for presenting risks and risk reductions. *Cochrane Database of Systematic Reviews*, *3*.

Ancker, J. S., Senathirajah, Y., Kikafka, R., & Starren, J. B. (2006). Design features of graphs in health risk communication: a systematic review. *Journal of the American Medical Informatics Association*, *13*, 608–618.

Arkes, H. R., & Gaissmaier, W. (2012). Psychological research and the prostate-cancer screening controversy. *Psychological Science*, *23*, 547–553.

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, *44*, 211–233.

Bastian, L. A., Nanda, K., Hasselblad, V., & Simel, D. L. (1998). Diagnostic efficiency of home pregnancy test kits: a meta-analysis. *Archives of Family Medicine*, *7*, 465–469.

Bodemer, N., Meder, B., & Gigerenzer, G. (2012). Communicating relative risk changes with baseline risk: the effect of presentation format and numeracy. Manuscript under revision.

Bramwell, R., West, H., & Salmon, P. (2006). Health professionals' and service users' interpretation of screening test results: experimental study. *British Medical Journal*, *333*, 284–289.

Brase, G. L. (2009). Pictorial representations and numerical representations in Bayesian reasoning. *Applied Cognitive Psychology*, *23*, 369–381.

Brainerd, C. J. (1981). Working memory and the developmental analysis of probability judgment. *Psychological Review*, *88*, 463–502.

Casscells, W., Schoenberger, A., & Graboys, T. B. (1978). Interpretation by physicians of clinical laboratory results. *The New England Journal of Medicine*, *299*, 999–1001.

Christensen-Szalanski, J. J. J., & Beach, L. R. (1982). Experience and the base-rate fallacy. *Organizational Behavior and Human Performance*, *29*, 270–278.

Christensen-Szalanski, J. J. J., & Bushyhead, J. B. (1981). Physicians' use of probabilistic information in a real clinical setting. *Journal of Experimental Psychology. Human Perception and Performance*, *7*, 928–935.

Cole, W. G., & Davidson, J. E. (1989). Graphic representation can lead to fast and accurate Bayesian reasoning. *Proceedings of the Annual Symposium on Computer Application in Medical Care*, *8*, 227–231.

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

Covey, J. (2007). A meta-analysis of the effects of presenting treatment benefits in different formats. *Medical Decision Making*, *27*, 638–654.

Daston, L. J. (1988). *Classical probability in the enlightenment*. Princeton: Princeton University Press.

Djulbegovic, M., Beyth, R. J., Neuberger, M. M., Stoffs, T. L., Vieweg, J., Djulbegovic, B., & Dahm, P. (2010). Screening for prostate cancer: systematic review and meta-analysis of randomized controlled trials. *British Medical Journal*, *341*, c4543.

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: heuristics and biases* (pp. 249–267). Cambridge: Cambridge University Press.

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment* (pp. 17–52). New York: Wiley.

Edwards, W., & von Winterfeldt, D. (1986). On cognitive illusions and their implications. In H. R. Arkes & K. R. Hammond (Eds.), *Judgment and decision making*. Cambridge: Cambridge University Press.

Edwards, A., Elwyn, G., Covey, J., Matthews, E., & Pill, R. (2001). Presenting risk information—a review of the effects of "framing" and other manipulations on patient outcomes. *Journal of Health Communication*, *6*, 61–82.

Edwards, A., Elwyn, G., & Mulley, A. (2002). Explaining risks: turning numerical data into meaningful pictures. *British Medical Journal*, *324*, 827–830.

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

Fagerlin, A., Wang, C., & Ubel, P. A. (2005). Reducing the influence of anecdotal reasoning on people's health care decisions: is a picture worth a thousand statistics? *Medical Decision Making*, *25*, 398–405.

Fischbein, E., Pampu, I., & Minzat, I. (1970). Comparison of ratios and the chance concept in children. *Child Development*, *41*, 377–389.

Galesic, M., García-Retamero, R., & Gigerenzer, G. (2009). Using icon arrays to communicate medical risks: overcoming low numeracy. *Health Psychology*, *28*, 210–216.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: a reply to Kahneman and Tversky. *Psychological Review*, *103*, 592–596.

Gigerenzer, G. (2002). *Calculated risks: how to know when numbers deceive you*. New York: Simon & Schuster.

Gigerenzer, G. (2007). *Gut feelings: the intelligence of the unconscious*. New York: Viking Press.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*, 451–482.

Gigerenzer, G. & Gray, J. A. M. (Eds.) (2011). *Better doctors, better patients, better decisions: envisioning health care 2020*. Cambridge: MIT Press.

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

Gigerenzer, G., & Hoffrage, U. (1999). Overcoming difficulties in Bayesian reasoning: a reply to Lewis and Keren (1999) and Meilers and McGraw (1999). *Psychological Review*, *106*, 425–430.

Gigerenzer, G., Hertwig, R., & Pachur, T. (Eds.) (2011). *Heuristics: the foundations of adaptive behavior*. New York: Oxford University Press.

Gigerenzer, G., Mata, J., & Frank, R. (2009). Public knowledge of benefits of breast and prostate cancer screening in Europe. *Journal of the National Cancer Institute*, *101*, 1216–1220.

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

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest*, *8*, 53–96.

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

Girotto, V., & Gonzalez, M. (2008). Children's understanding of posterior probability. *Cognition*, *106*, 325–344.

Hammerton, M. (1973). A case of radical probability estimation. *Journal of Experimental Psychology*, *101*, 252–254.

Hertwig, R., Hoffrage, U., & the ABC Research Group (2012). *Simple heuristics in a social world*. New York: Oxford University Press.

Hoffrage, U., & Gigerenzer, G. (1998). Using natural frequencies to improve diagnostic inferences. *Academic Medicine*, *73*, 538–540.

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

Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Communicating statistical information. *Science*, *290*, 2261–2262.

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M. S., & Caverni, J.-P. (1999). Naive probability: a mental model theory of extensional reasoning. *Psychological Review*, *106*, 62–88.

Kahneman, D., & Tversky, A. (1972). Subjective probability: a judgment of representativeness. *Cognitive Psychology*, *3*, 430–454.

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

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, *103*, 582–591.

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

Kaye, D. H., & Koehler, J. J. (1991). Can jurors understand probabilistic evidence? *Journal of the Royal Statistical Society. Series A*, *154*, 75–81.

Kleiter, G. (1994). Natural sampling: rationality without base rates. In G. H. Fischer & D. Lang (Eds.), *Contributions to mathematical psychology, psychometrics, and methodology* (pp. 375–388). New York: Springer.

Knight, F. H. (1921/2006). *Risk, uncertainty and profit*. Boston: Houghton Mifflin.

Koehler, J. J. (1996a). The base rate fallacy reconsidered: descriptive, normative and methodological challenges. *Behavioral and Brain Sciences*, *19*, 1–54.

Koehler, J. J. (1996b). On conveying the probative value of DNA evidence: frequencies, likelihood ratios and error rates. *University of Colorado Law Review*, *67*, 859–886.

Koehler, J. J. (2006). Train our jurors. In G. Gigerenzer & C. Engel (Eds.), *Heuristics and the law* (pp. 303–326). Cambridge: MIT Press.

Kurz-Milcke, E., Gigerenzer, G., & Martignon, L. (2008). Transparency in risk communication: graphical and analog tools. *Annals of the New York Academy of Sciences*, *1128*, 18–28.

Kurz-Milcke, E., & Martignon, L. (2006). Lebendige Urnen und ereignisreiche Bäume: Überlegungen und Versuche zu einer Didaktik der Stochastik in der Grundschule [Lively urns and eventful trees: developing a didactics of stochastics in elementary school]. *Anregungen zum Stochastikunterricht*, *3*, 182–203.

Kurzenhäuser, S., & Hoffrage, U. (2002). Teaching Bayesian reasoning: an evaluation of a classroom tutorial for medical students. *Medical Teacher*, *24*, 516–521.

Labarge, A. S., McCaffrey, R. J., & Brown, T. A. (2003). Neuropsychologists' abilities to determine the predictive value of diagnostic tests. *Archives of Clinical Neuropsychology*, *18*, 165–175.

Lewis, C., & Keren, G. (1999). On the difficulties underlying Bayesian reasoning: comment on Gigerenzer and Hoffrage. *Psychological Review*, *106*, 411–416.

Lindsey, S., Hertwig, R., & Gigerenzer, G. (2003). Communicating statistical DNA evidence. *Jurimetrics*, *43*, 147–163.

Lipkus, I. M., & Hollands, J. G. (1999). The visual communication of risk. *Journal of the National Cancer Institute Monographs*, *25*, 149–163.

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, *21*, 37–44.

Martigon, L. (2013). Fostering children's probabilistic reasoning and their risk evaluation. In E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic thinking: presenting plural perspectives*. Dordrecht: Springer. This volume.

Martignon, L., & Krauss, S. (2009). Hands on activities with fourth-graders: a tool box of heuristics for decision making and reckoning with risk. *International Electronic Journal for Mathematics Education*, *4*, 117–148.

Martignon, L., & Krauss, S. (2007). Gezinkte Würfel, Magnetplättchen und Tinker Cubes für eine Stochastik zum Anfassen in der Grundschule. *Stochastik in der Schule*, *27*, 21–30.

Meder, B., & Nelson, J. D. (2012). Information search with situation-specific reward functions. *Judgment and Decision Making*, *7*, 119–148.

Meder, B., Mayrhofer, R., & Waldmann, M. R. (2009). A rational model of elemental diagnostic inference. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st annual conference of the Cognitive Science Society* (pp. 2176–2181). Austin: Cognitive Science Society.

Moyer, V. A., on behalf of the U.S. Preventive Services Task Force (2012). Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement. *Annals of Internal Medicine*, *157*, 120–134.

Multmeier, J. (2012). *Representations facilitate Bayesian reasoning: computational facilitation and ecological design revisited*. Unpublished doctoral dissertation, Free University, Berlin.

Nelson, J. D., McKenzie, C. R. M., Cottrell, G. W., & Sejnowski, T. J. (2010). Experience matters: information acquisition optimizes probability gain. *Psychological Science*, *21*, 960–969.

Peters, E. (2008). Numeracy and the perception and communication of risk. *Annals of the New York Academy of Sciences*, *1128*, 234–267.

Peterson, C. R., & Beach, L. R. (1967). Man as intuitive statistician. *Psychological Bulletin*, *68*, 29–46.

Peterson, C. R., & Miller, A. J. (1965). Sensitivity of subjective probability revision. *Journal of Experimental Psychology*, *70*, 117–121.

Peterson, C. R., Schneider, R. J., & Miller, A. J. (1965). Sample size and the revision of subjective probabilities. *Journal of Experimental Psychology*, *69*, 522–527.

Phillips, L. D., & Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of Experimental Psychology*, *72*, 346–354.

Piaget, J., & Inhelder, B. (1975). *The origin of the idea of chance in children*. New York: Norton. (Original work published 1951).

Rouanet, H. (1961). Études de décisions expérimentales et calcul de probabilités [Studies of experimental decision making and the probability calculus]. In *Colloques Internationaux du Centre National de la Recherche Scientifique* (pp. 33–43). Paris: Éditions du Centre National de la Recherche Scientifique.

Schwartz, L., Woloshin, S., Black, W. C., & Welch, H. G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine*, *127*, 966–972.

Sedlmeier, P. (1999). *Improving statistical reasoning: theoretical models and practical implications*. Mahwah: Erlbaum.

Sedlmeier, P., & Gigerenzer, G. (2001). Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology. General*, *130*, 380–400.

Simon, H. (1969). *The science of the artificial*. Cambridge: MIT Press.

Steurer, J., Fischer, J. E., Bachmann, L. M., Koller, M., & ter Riet, G. (2002). Communicating accuracy of tests to general practitioners: a controlled study. *British Medical Journal*, *324*, 824–826.

Thompson, W. C., & Schumann, E. L. (1987). Interpretation of statistical evidence in criminal trials. *Law and Human Behavior*, *11*, 167–187.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, *185*, 1124–1131.

Wegwarth, O., Schwartz, L. M., Woloshin, S., Gaissmaier, W., & Gigerenzer, G. (2012). Do physicians understand cancer screening statistics? A national survey of primary care physicians. *Annals of Internal Medicine*, *156*, 340–349.

Wegwarth, O., Gaissmaier, W., & Gigerenzer, G. (2011). Deceiving numbers: survival rates and their impact on doctors' risk communication. *Medical Decision Making*, *31*, 386–394.

Weir, B. S. (2007). The rarity of DNA profiles. *Annals of Applied Statistics*, *1*, 358–370.

Yost, P. A., Siegel, A. E., & Andrews, J. M. (1962). Non-verbal probability judgments by young children. *Child Development*, *33*, 769–780.

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

Fostering Children's Probabilistic Reasoning and First Elements of Risk Evaluation

Laura Martignon

Abstract The ABC of probabilistic literacy for good decision-making should be conveyed to children at an early stage, or more precisely, before they reach their 11th year of age. This conviction is based on the view of experts who maintain that the mathematical competencies of adults who are not especially trained in mathematical subjects are those they developed when they were 9, 10 and 11 years old. This paper will show how children can be provided with elementary, yet useful tools for decision making under uncertainty. Children, as has been demonstrated empirically, can acquire a mosaic of simple, play-based activities, by means of tinker-cubes. The results reported here were guided and inspired by empirical studies on human decision making obtained by the Centre of Adaptive Behaviour and Cognition, directed by Gerd Gigerenzer.

1 Introduction

In the decade of 1960s, two scientists, married to each other, caused a paradigm stir not just in linguistics, their own field, but in cognitive sciences with a wide range of consequences, even for a seemingly distant field of Mathematics Education. These scientists were Carol and Noam Chomsky and the question they set out to answer was whether humans are innate grammarians. It has become a popular misconception that Noam Chomsky proved language is innate, a view that has been perpetuated by scientists. Chomsky never proved any such thing. He did postulate, however, that while both a human baby and a kitten are capable of some form of inductive reasoning, the human child will eventually acquire the competency to understand and produce language, while the kitten will never acquire either competency. Carol Chomsky's empirical studies showed that children do improve their language competency, but this improvement is achieved through learning and instruction. From the seminal findings of these two scientists, two research directions have originated.

L. Martignon (✉)
Institute of Mathematics and Computer Science, Ludwigsburg University of Education, Reuteallee 46, 71634 Ludwigsburg, Germany
e-mail: laura_martignon@yahoo.it

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_9,

On the one hand, researchers look for innate dispositions for a variety of competencies and, on the other hand, researchers look for dispositions for "acquiring" specific competencies "through instruction." Thus, questions like: "Are humans intuitive logicians?" or, "Do humans have the disposition for acquiring logical competencies through instruction?" have guided prolific research with important consequences for education.

In 1969, Amos Tversky and Daniel Kahneman began analyzing humans' innate ability for statistical reasoning (Kahneman 2011) guided by the question: Are humans intuitive statisticians?

As Kahneman narrates in his recent book "Thinking fast and slow" (Kahneman 2011), this question was one of the topics proposed by Amos Tversky for a seminar at the Psychology Department of the University of Jerusalem. This was the starting point of a research program whose results changed the basis of cognitive science with important consequences for probabilistic education. Tversky and Kahhneman coined the expression 'heuristics and biases' as title of their program, and were responsible for a paradigm shift in cognitive psychology. The thesis underlying the program, based on a flurry of impressive empirical studies, was that human thought is not necessarily probabilistic, whereas it often relies on heuristics for reasoning and decision making. They discovered a number of important biases that tend to riddle probabilistic thought. The list of these biases by Tversky and Kahneman and by the school they initiated is long. We will restrict ourselves to the two that are most significant for education in probability reasoning.

Base rate neglect in Bayesian reasoning: In making inferences about probabilities, decision makers tend to ignore the background frequencies. For example, if the probability of any given woman having breast cancer is known to be $1/10{,}000$—the base rate—but a test on 10,000 women gives 100 positive results, decision makers will tend to overestimate the probability that any one of the women testing positive actually has cancer, and imagine it of the same order of magnitude as the sensitivity of the test.

Risk-aversion: This is one of the foundations of the heuristics and biases program that has challenged the economic "rational actor" model of intuitive human reasoning. Risk-aversion refers to the fact that losses loom larger than gains. An individual may take great pains to avoid small losses, while neglecting strategies to maximize long-term gains.

A new school of thought, led by Gerd Gigerenzer, proposes that at least to some extent, the negative results on biases in probabilistic reasoning can be explained by discrepancies between the formats of information and the tasks on which humans perform so poorly today, and those faced by our ancient forebears (e.g., Gigerenzer et al. 1999). Gigerenzer's thesis (Gigerenzer and Hoffrage 1995) is that the human mind is well adapted to frequencies that are sampled "naturally", whereas it functions poorly when confronted with Kolmogorov probabilities or with percentages. This research has been empirically proved by him and his co workers.

The question is then: *Would it be possible to improve the public's skill at probabilistic reasoning by matching pedagogical strategies adaptively to cognitive processes during early phases of education*?

Doing this would contribute to providing anchoring mechanisms and "translation" heuristics for the phase when more formal representations are usually taught.

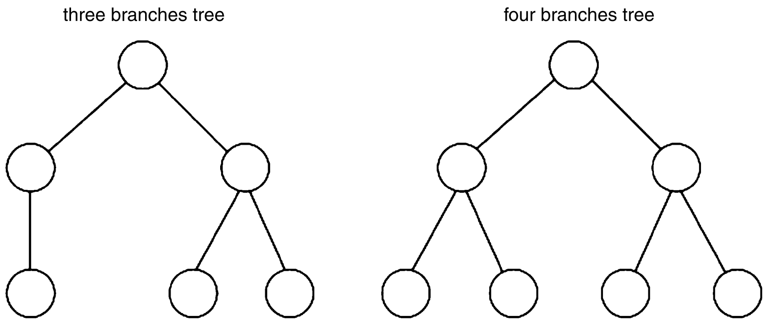
This paper illustrates the first steps in the direction of the question posed above. In this paper, we build on existing results concerning the cognitive mechanisms underlying probabilistic reasoning.

2 Result Verification in the Bayesian Reasoning Task

The prototypical Bayesian reasoning task involves using evidence about an uncertain proposition to revise an assessment of the likelihood of a related proposition. The following example is drawn from an article by Zhu and Gigerenzer (2006) that examined children's ability to perform Bayesian reasoning. The context was a small village in which "red nose" was a "symptom" of "telling lies." Children needed to relate the proposition "having a red nose" to the proposition "telling lies." Specifically, they were given information about the probability of "red nose" conditioned on "liar" and that of "red nose" conditioned on "non-liar." They were also informed about the incidence of liars in the village. They were then asked to establish the chance that someone with a red nose tells lies. The performance of children (fourth graders) improved substantially when the probabilistic setting was replaced by a setting in which the cover story reported the natural frequencies involved, i.e., in terms of a sequential partitioning of nested sets and their proportions. Humans, children and adults, appear to be adapted to a "natural" sequential partitioning for categorization. The term "natural" means that these proportions (i.e., relative frequencies) are perceived as being obtained by the mental simulation of counting. Atmaca and Martignon (2004) conjectured that different neural circuits are involved in the natural frequency and probability versions of the Bayesian task and reported experimental results that support their conjecture. Subjects were given tasks using a slide projector, and solved them as mental arithmetic, without writing. Data about correctness of solutions and the time to solution was collected. The experiment made use of a response mode called *result verification* or *result disparity* (Kiefer and Dehaene 1997): Subjects are presented with a proposed solution and asked to judge as quickly as possible whether it is correct or incorrect. Atmaca and Martignon found that subjects needed significantly longer times and produced significantly fewer correct answers, for the tasks given in probability format versus those given in the natural frequency format. In one experiment, 110 participants were exposed to typical Bayesian tasks with "three branches" and tasks with "four branches", see Fig. 1.

The results of the experiment are summarized in Fig. 2.

Perception of frequencies of occurrences, this result suggests, could be a mechanism, or at least part of a more complex mechanism, that enables fast and effective decisions in uncertain situations, because "…*natural selection* (…) *gives rise to practical cognitive mechanisms that can solve* (…) *real world problems…*" (Fiddick and Barrett 2001, p. 4). Dehaene (1997) wrote in a similar context: "*Evolution has been able to conceive such complex strategies for food gathering, storing, and*



Three branches: 10 out of 1000 children have German measles. Out of the 10 children who have German measles, all 10 have a red rash. *Of the 990 children without German measles, 9 also have a red rash. How many of the children with a red rash have the German measles*?
Four branches: *10 out of 1000 car drivers have an accident at night. Out of the 10 car drivers who have an accident at night, 8 are intoxicated. Out of the 990 car drivers who do not have an accident at night, 40 also are intoxicated. How many of the car drivers who are intoxicated actually have an accident at night*?

Fig. 1 Bayesian task in two modalities

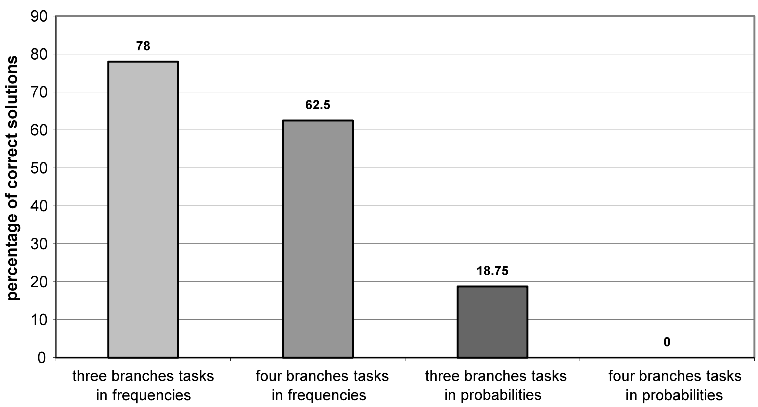


Fig. 2 Percentage of subjects solving the Bayesian task in two modalities correctly

predation, that it should not be astonishing that an operation as simple as the comparison of two quantities is available to so many species)." (Dehaene 1997, p. 27)

3 The Automatic Processing of Frequencies

The automatic perception of frequencies of occurrences was described by Hasher and Zacks in the 1970s: "Operations that drain minimal energy from our limited-capacity attention mechanism are called automatic; their occurrence does not interfere with other ongoing cognitive activity. They occur without intention and do not benefit from practice." We propose that humans are genetically 'prepared' for certain automatic processes: These processes encode the fundamental aspects of the flow of information, namely, spatial, temporal, and frequency-of-occurrence information" (Hasher and Zacks 1979, p. 356). Humans are known to be sensitive to frequencies even when they do not pay attention to them. In experiments, participants performed well when remembering frequencies of events, even when unprepared for the memory test (Zacks et al. 1982). There is also evidence to suggest that they did not count the events (Coren and Porac 1977). As reported by Hasher and Zacks (1984), experiments have shown that the processing of frequencies of occurrence information is neither activated nor influenced by age, training, feedback, individual differences such as intelligence, knowledge or motivation, nor reductions in cognitive capacity such as depression or multiple task demand. In other words, there is some suggestion here that the human brain adapted to the processing of frequencies during evolution. Furthermore, several experiments have provided evidence that "(...) various animal species including rats, pigeons, raccoons, dolphins, parrots, monkeys and chimpanzees can discriminate the numerosity of various sets, including visual objects presented simultaneously or sequentially and auditory sequences of sounds" (Dehaene et al. 1998, p. 357). According to models of animal counting presented by Meck and Church (1983), numbers are represented internally by the continuous states of an analogue accumulator. For each counted item, a more-or-less fixed quantity is added to the accumulator. The final state of the accumulator therefore correlates well with numerosity, although it may not be a completely precise representation of it (Dehaene 1992). This model explains the observation that animals are very good in handling small quantities, while performance degrades with the increase of magnitude. During this counting process, "the current content of the accumulator is used as a representative of the numerosity of the set so far counted in the decision processes that involve comparing a current count to a remembered count" (Gallistel and Gelman 1992, p. 52). Thus the current content of the accumulator represents the magnitude of the current experienced numerosity, whereas previously read out magnitudes are represented in long-term memory. This enables comparison of two number quantities which is essential for frequency processing. The comparison of number processing abilities in animals and human infants leads to the conclusion "that animal number processing reflects the operation of a dedicated, biologically determined neural system that humans also share and which is fundamental to the uniquely human ability to develop higher-level arithmetic" (Dehaene et al. 1998, p. 358). Several studies report abilities of frequency perception in kindergarten and elementary school children in grades 1 to 6 (Hasher and Zacks 1979; Hasher and Chromiak 1977). And the findings of abilities in numerosity discrimination in infants and even newborns (Antell and Keating 1983) "may indicate that some capacity for encoding frequency is present from birth" (Hasher and Zacks 1984, p. 1378).

4 From Absolute Numerical Quantities to Natural Frequencies

Animals and humans could not survive if they had developed only a single, simple sense for absolute frequencies without a sense for *proportions of numerical quantities* for inference. “Are all red mushrooms poisonous, or only some of them? How valid is the color red as an indicator of danger from poisoning for mushrooms?” Whereas non-precise estimates may have been sufficient for survival in ancient rural societies, in modern human communities answering this type of question using well calibrated inferences is vital. Successful citizenry requires this type of competency. We are convinced that elementary school should provide tools for successful quantified inferences. In order to establish whether one cue is a better predictor than another (e.g., whether the color red is a better predictor than white dots for poisonous mushrooms), we need a well tuned mental mechanism that compares proportions. Within the realm of cognitive neuroscience, little is known about the brain processes that are involved in proportion estimation and proportion comparison. The more rudimentary instruments for approximate inference that humans share with animals must be based on some sort of non-precise proportion comparison. But when do infants in modern societies begin to quantify their categorization and when can they be trained in quantified proportional thinking? Although fractions are the mathematical tool for describing proportions, we envisage an early preparation of children’s use of numerical proportions without “normalizations” even before they are confronted with fractions. The results by Piaget and Inhelder on the understanding of such proportions in children motivated two generations of research in developmental and pedagogical psychology, specifically in the learning of mathematics. We cite here one direction in particular which has been fundamental to our work (Koerber 2003). In a series of well designed experiments, Stern et al. (2002) demonstrated that third-graders can *learn to abandon* the so-called *additive misconception* in which children respond with “9” instead of “12” to “3 : 6 = 6 :?”. Children were asked to compare mixtures of lemons and oranges in terms of their intensity of taste. The training used a balance beam or graphs to represent juice mixtures with the movement of a pivot representing changes in proportions. At the end of a short training (two days maximum), children showed improvement in proportional thinking, thus providing evidence of third-graders’ aptitude to learn proportional thinking when provided with adequate instruction.

5 Cognitively Natural Representations and Task Performance

Stern’s results can be combined with Gigerenzer’s at the interface between pedagogy and cognitive science. He and his colleagues searched for pedagogical approaches that tap into cognitively natural representations. The *natural frequency* representation for Bayesian reasoning tasks is based on information that can be gained by “naturally” counting events in an environment, and therefore taps into very basic human information processing capacities. “*Natural sampling is the way* *humans have encountered statistical information during most of their history. Collecting data in this way results in natural frequencies*" (Hoffrage et al. 2002).

The term "natural" signifies that these frequencies have not been normalized with respect to base rates. Probabilities and percentages can be derived from natural frequencies by normalizing natural frequencies into the interval: [0, 1] or [0, 100], respectively. However, this transformed representation results in loss of information about base rates.

Consider the following examples from Hoffrage et al. (2002):

> Natural frequencies: *Out of each* 100 *patients,* 4 *are infected. Out of* 4 *infected patients,* 3 *will test positive. Out of* 96 *uninfected patients,* 12 *will also test positive*.
> Normalized frequencies: *Out of each* 100 *patients,* 4 *are infected. Out of* 100 *infected patients,* 75 *will test positive. Out of* 100 *uninfected patients,* 12.5 *will also test positive*.

Because normalized frequencies filter out base rate information, they make the Bayesian task of inferring a posterior probability from evidence more difficult. The inference that most positives are false positives can be read directly from the natural frequency representation, while it must be obtained via a non-trivial calculation from the normalized frequency representation. We claim that training young children in an interactive way with natural proportions enables them to make use of simple heuristics for dealing with probabilities. We worked with fourth graders, preparing them to solve one of the mathematical test items of PISA 2003 namely: *Consider two boxes A and B. Box A contains three marbles, of which one is white and two are black. Box B contains* 7 *marbles, of which two are white and five are black. You have to draw a marble form one of the boxes with your eyes covered. From which box should you draw if you want a white marble*? Only 27 % of the German school students were able to justify that one should choose Box A. Serious amounts of conceptual work are needed for mathematically correct statements regarding why and when larger proportions in samples correspond to larger chances in the populations. In early grades, the consensus is that one should focus on intuition and competency rather than on formal mathematics. We should provide students with (i) basic stochastic modeling skills with natural representation formats and (ii) simple heuristics for operating with these formats. With this hypothesis in mind, Martignon and Kurz-Milcke (2005) designed a program that develops and encourages the natural frequency representation through the use of enactive learning (see also Kurz-Milcke and Martignon 2005). In playful, yet structured activities, children use colored plastic cubes called tinker-cubes to represent individuals that make up a population. Different colors represent different attributes (e.g., red cubes for girls; blue for boys). The cubes can be attached one to another, allowing representation and multi-attribute encoding (e.g., a red cube attached to a yellow cube for a girl with glasses; a blue cube attached to a green cube for a boy without glasses, a red cube attached to a green cube for a girl without glasses). Children collect the tinker-cubes into plastic urns that represent populations this gaining concrete, visual and tactile experience with individuals with multiple attribute combinations and their grouping into categories and subcategories. Recent exploratory studies of fourth-grade children indicate that children are both enthusiastic and successful when constructing these representations of categories and sub-categories in nested set (see Fig. 3

Fig. 3 Constructing our class

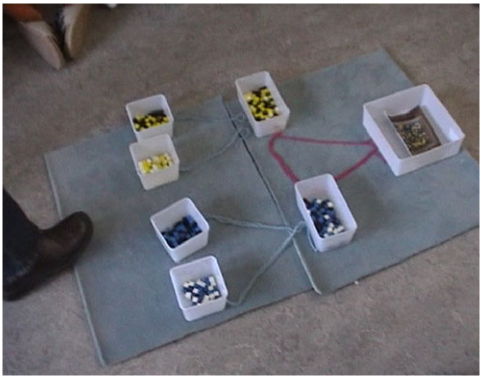


Fig. 4 Enactive Bayesian reasoning 4th—graders (Kurz-Milcke and Martignon 2005), photos by Kurz-Milcke

and Fig. 4), especially when the populations are personally meaningful (e.g., "our class"). They can easily "construct" answers to questions like: "How many of the children wearing glasses are boys?"

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

6 Fourth-graders can reckon with risk[1]

Reckoning with risk has become all the more subtle for humans in modern times since most risks are communicated by means of mathematical formats that have to be learned in school. Both medical and finance decision making is based on statistical information. Thus, training young students in the perception of risk has become fundamental in modern society. Students have to acquire tools to model the risk of having a disease like AIDS, or of being pregnant when the pregnancy test is positive, of losing money with an investment, etc. Expressed in mathematical terms, a risky event is one associated with a strictly positive probability of a loss of resources like health, time, food, or money. These resources are usually modeled in terms of "utilities." The framework for understanding risk can be taught in secondary school, but the basic intuitions can—and should be—fostered already in primary school. We argue that by reducing probabilities to proportions a good understanding of risks can be fostered. Stochastic activities in primary school cannot be limited to urns, albeit these are the basis of a good stochastic training. Young students have to learn to reflect on resources and on loss of resources. There are many popular board games for children involving possible losses and gains of resources. These games can be used to illustrate risky behavior. For instance, German children begin playing Ludo (German "Mensch ärgere Dich nicht") when they are five years old. We give a short account of the game's instructions:

> Ludo: Players take it in turn to throw a single die. A player must first throw a six to be able to move a piece from the starting area onto the starting square. In each subsequent turn, the player moves a piece forward 1 to 6 squares as indicated by the die. When a player throws a 6 the player may bring a new piece onto the starting square, or may choose to move a piece already in play. Any throw of a six results in another turn. If a player cannot make a valid move, he/she must pass the die to the next player. If a player's piece lands on a square containing an opponent's piece, the opponent's piece is captured and returns to the starting area. A piece may not land on square that already contains a piece of the same color (unless playing doubling rules; see below). Once a piece has completed a circuit of the board it moves up the home column of its own color. The player must throw the exact number to advance to the home square. The winner is the first player to get all four of his/her pieces onto the home square.

We proposed the following task in 6 classes at grade 4 at elementary schools in and near Stuttgart:

> *Consider two players "Blue" and "Green." In one position, Blue is two squares behind Green, and at another position, Blue is three squares behind Green. Assume it is Green's turn. He draws a "one." Which token should Green move? Which move is riskier?*

We asked each of 193 children, in which of the two green tokens he/she would move if he/she were Green and rolled a "one." We were not just interested in the answers but in the discussion on the risk associated with each of the two possible

[1] Part of this section is derived from Martignon and Krauss (2009). Hands on activities with fourth-graders: a tool box of heuristics for decision making and reckoning with risk. International Electronic Journal for Mathematics Education, vol. 4, Number 3, 117–148.

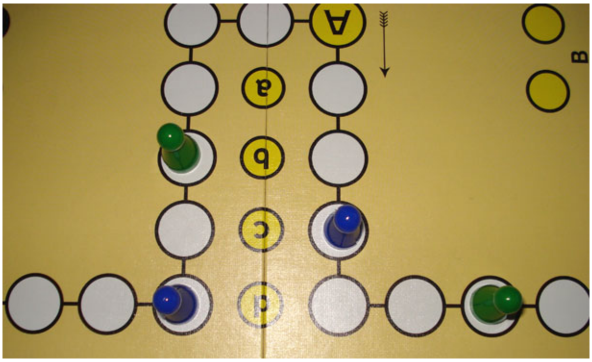


Fig. 5 A situation in the game Ludo

moves. Note that when the left green token (see Fig. 5) moves "one" step forward, blue can only chase a green token away in his next move after rolling a "3" (which will be the situation in 1 out of 6 possible cases). If the right green token moves one step forward, blue can chase one green token away by either rolling a "2" or "4" (which will be in 2 out of 6 cases). Therefore, in the first option, the chances of having one token chased away are "1 out of 6" while in the second they are "1 out of 3" (2 out of 6). Only 21 of the 193 children wrote that they would choose to move the left green token and gave a conclusive explanation. It seems that children and adults as well would tend to move forward with the first green token, ignoring the higher risk of losing a token at the next move of the adversary.

Consider the situation described above, where the risk of losing one token at Ludo can have a probability of "one out of three" if one moves further with the right token (Fig. 5) or "one out of six" if one moves with the left token. How might one compare the risk in the two scenarios? Would it be correct to say, that "*the expectancy of Green keeping a token has been increased by 100 %*," when actually "*the risk of losing one token has been reduced from* "*two out of six*" *to* "*one out of six*" "?

The 193 children were instructed on the meaning of the reduction from "1 out of 3" to "1 out of 6" and were then tested in other similar situations in Ludo.

This type of instruction may seem trivial, yet it is central to the understanding of risks in the way they are communicated in modern media. Both in the context of finance and medicine, "relative" risks are often presented instead of "absolute" risks, thus blurring the understanding of their real impact. In the above example, the risk of losing a token can be increased by 100 % (because 2 are 100 % more than 1) if one moves the right hand token. Yet, relatively to the 6 possible events, the relative risk of losing a token can be reduced by 1/6 (i.e., about 17 %) by moving the left token.

Let us look at a typical example of the same phenomenon which is highly relevant in adult life: Women are constantly prompted to perform regular screening (mammography) to enhance their life expectancy by 25 %. How should this enhancement be interpreted? Gigerenzer (2004) has analyzed this frequent type of communication format in great detail. The 25 % enhancement of life expectancy can be explained as follows: Four out of 1000 women have breast cancer, and if these 1000 women perform screening regularly then one will be saved. It is crucial to become aware that the corresponding percentage would be 0.1 % (1/1000) when communicated as relative risk. Gigerenzer (2004) has looked at many other examples and the dramatic consequences of this type of confusion between "relative" and "absolute" risk.

The Ludo game situations presented above can be a first encounter of fourth graders with the important difference between absolute and relative risks.

The interactive site www.eeps.com/riskicon has been created by us for teaching not just children but also future teachers how to deal with Bayesian reasoning with "natural" frequencies and also with relative and absolute risk reduction. It is a dynamic site that allows simulations with varying parameters. So far this site has been used successfully with a group of future teachers of an international school in Berlin and with school students of a ninth class in this international school. We are presently planning a teachers' training by means of dynamic illustrations in this site.

Acknowledgements The author thanks Lynn Michell for useful suggestions both on aspects of this chapter and on the formulations in the text. This work has been sponsored by the SPP 15-16 of the DFG, in the Project Models of Information Search.

References

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

Antell, S. E., & Keating, D. P. (1983). Perception of numerical invariance in neonates. *Child Development*, *54*, 695–701.

Coren, S., & Porac, C. (1977). Fifty centuries of right-handedness: the historical record. *Science*, *198*, 631–632.

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

Dehaene, S. (1997). *The number sense*. London: Oxford University Press.

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neurosciences*, *21*(8), 355–361.

Fiddick, L., & Barrett, H. C. (2001). Evolution of cognition: an adaptationist perspective. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (Vol. 7, pp. 4996–5000). Amsterdam: Elsevier.

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, *44*, 43–74.

Gigerenzer, G. (2004). *Calculated risks: what numbers really tell us about our lives*. New York: Simon and Schuster.

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

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

Hasher, L., & Chromiak, W. (1977). The processing of frequency information: an automatic mechanism? *Journal of Verbal Learning and Verbal Behavior*, *16*, 173–184.

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

Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information—the case of frequency of occurrence. *The American Psychologist*, *39*(12), 1372–1388.

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

Kahneman, D. (2011). *Thinking fast and slow*. New York: Farrar, Straus and Giroux.

Kiefer, M., & Dehaene, S. (1997). The time course of parietal activation in single-digit multiplication: evidence from event-related potentials. *Mathematical Cognition*, *3*(1), 1–30.

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

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

Martignon, L., & Krauss, S. (2009). Hands on activities with fourth-graders: a tool box of heuristics for decision making and reckoning with risk. *International Electronic Journal for Mathematics Education*, *4*(3), 117–148.

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

Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes*, *9*(3), 320–334.

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

Zacks, R. T., Hasher, L., & Sanft, H. (1982). Automatic encoding of event frequencies—further findings. *Journal of Experimental Psychology—Human Learning and Memory*, *8*(2), 106–116.

Zhu, L., & Gigerenzer, G. (2006). Children can solve Bayesian problems: The role of representation in mental computation.

Intuitive Conceptions of Probability and the Development of Basic Math Skills

Gary L. Brase, Sherri Martinie, and Carlos Castillo-Garsow

Abstract The idea of probabilities has been described as a "Janus-faced" concept, which can be thought of either in terms of frequencies or in terms of subjective confidence. This dualism contributes to debates about the nature of human rationality, and therefore the pedagogical assumptions and goals of education. For this reason, the present chapter explores the evidence regarding how quantitative information is intuitively understood in the human mind over the course of elementary school education. Are particular interpretations of probability equally weighted or does one interpretation predominate as mathematical concepts are being acquired? We find multiple, converging lines of evidence that indicate a frequency interpretation of probabilistic information is developmentally primary and privileged. This has implications for mathematics education, even before the introduction of actual probabilities, in areas such as learning fractions and decimals. Educational practices should work to bootstrap from these privileged representations (rather than fight them) and built towards a more inclusive and comprehensive model of probability knowledge. We conclude that a fundamental issue is not just whether students think about probabilities as a frequentist or as a subjectivist, but rather how they recognize when to be one versus the other.

1 The Interpretation of Probabilities

Mathematics may be the queen of the sciences, but this queen has a dirty little secret. Once in a while, she sneaks out and cavorts with the common people; arguing, gambling, and getting into fights. When she does this it is called *probabilities*. Far from the regal courtyards, the mathematics of probabilities developed as a way to evaluate what were good or bad gambles (Gigerenzer et al. 1989). She engaged in ferocious debates about what she was and how she should behave (Gigerenzer 1993). She behaves, in other words, in a very un-queenly manner.

In all seriousness, the concept of probabilities in mathematics really does lead a metaphorical double life. On the one hand, there is a conception of probabil-

G.L. Brase (✉) · S. Martinie · C. Castillo-Garsow
Kansas State University, Manhattan, KS, USA
e-mail: gbrase@ksu.edu

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_10,

ities that involves using past events to statistically predict future outcomes. For example, a "30 % chance of rain" is based on a model of past weather events (It has rained on 3 out of the 10 previous days similar to this one). We will call this the *frequentist* concept of probabilities because it is, at its core, based on frequencies of past events. On the other hand, there is a conception of probabilities that involves degrees of belief. For example, a "1 % chance that the earth will be destroyed in 10 years" is based on a person's subjective beliefs, not a history of past events. We will call this the *subjectivist* concept of probabilities because it is, at its core, based on subjective belief states. Both concepts of probabilities can be legitimately claimed as correct, which led Hacking (1975) to refer to probability as a "Janus-faced" concept. These two conceptions of probability are sometimes easily distinguished, but at other times the distinction can be very subtle. For instance, a frequentist could claim that the probability of a six-sided die showing an even number is 0.5, based on a history of past rolls (i.e., of the prior rolls, half of them came up even). On the other hand, a subjectivist could claim the same answer by counting the sides, dividing by the number of sides with even numbers, assuming the die is fair, and predicting (without ever rolling the die once) that the probability of an even roll is 0.5.

The polysemous state of the probability concept is not constrained as just a mathematical annoyance. It also connects with disagreements about the nature of human rationality. After all, having two different conceptions of probability can lead to two people having different answers to the same question yet both believing they are rational and correct. This type of discrepancy fuels a debate about the nature of rationality in psychology research (see also the chapter by Ejersbo and Leron).

On one side of this debate about rationality is the *ecological rationality* view, which hews more closely to the frequentist conception of probability. This view is a more recent approach in psychology, although it has an intellectual history as long or longer than the alternative view (Gigerenzer et al. 1989). The ecological rationality view stresses the fit between the structure of the natural environment, which includes a frequentist focus on encountered objects, events, and locations, and the structure of the human mind. The emphasis here is on the rationality of making the best choice *under ecologically realistic circumstances*. Specifically, what is the best choice to make given the time and computational constraints of a situation and taking advantage of the inherent statistical properties of the natural world within which the choice is being made? (see Gigerenzer et al. 1999; also see chapters by Meder and Gigerenzer and by Martignon).

The other side of this debate about rationality is the *heuristics and biases* view, which is not as strongly affiliated with a particular view of probability but leans more towards the subjectivist conception. This view is more traditional in psychology, and in many ways is the current default approach, with seminal works by Tversky and Kahneman that go back to the 1970s (Kahneman and Tversky 1973; Tversky and Kahneman 1971, 1973, 1983). The heuristics and biases view emphasizes normative criteria for rationality, such that there is *a* correct answer and that answer is often (but not always) based on the subjectivist conception of probability. More recently, this heuristics and biases view has been shifting towards a dual process approach, in which people are thought to employ either deliberative styles of thought or intuitive styles of thought when reasoning (Kahneman 2011).

Why is this situation important for mathematics education? Because any approach to education must take into account what the learner is bringing with them to the educational endeavor. In very simple terms, the ecological rationality view describes people as intuitive frequentists, whereas the heuristics and biases view describes people as somewhat muddled subjectivists. Effective teaching and learning needs to take into account what the learner already knows (and hence does not need to be taught), what needs to be taught, and what should not be taught (either because it is not the appropriate point in time or because it is inherently confusing or wrong). An accurate employment of these consideration depends on what the nature of the human mind is regarding mathematical thinking.

2 Overview

The bulk of this chapter addresses a basic question, derived from the ecological rationality approach as contrasted with the heuristics and biases approach: *Are frequentist interpretations of probability privileged in the human mind?* Whereas the heuristics and biases approach has been very well documented as an approach to the learning and teaching of probabilities (e.g., Jones and Thornton 2005), this more recent ecological rationality approach is less well understood and appreciated. The basic question of this chapter is, in many ways, the central question regarding how these approaches hold implications for mathematics education. The heuristics and biases approach, in contrast to the ecological rationality approach, contends that both frequentist and subjectivists interpretations of probability are equally weighted as representations that the human mind can use for understanding probabilities. And, in fact, the subjectivist interpretations is often considered superior and thus to be promoted directly. These foundational assumptions color the pedagogical background assumptions, educational approaches, and expected learning outcomes of mathematics education. This chapter will therefore review the evidence that the ecological rationality approach, with its position that frequentist representations are cognitively privileged, is at least as valid as the heuristics and biases position.

We will use the method of converging operations to assess the evidence regarding this question (Garner et al. 1956; Sternberg and Grigorenko 2001a, 2001b). In short, the following sections will look at different areas of research and assess what the balance of evidence within each area says about our question. (See also Schmitt and Pilcher 2004 for a similar approach.) Specifically, we will be looking at evidence from evolutionary biology, information theory, phylogenetic research, cognitive psychology, developmental psychology, cross-cultural research, and neuroscience.

3 What Is the Evolutionary "History" of Quantitative Information?

Information from the natural environment is primarily frequencies—numbers of objects, occurrences of events, visits to locations—which can be tabulated in a binary fashion that produces basic, whole-number frequencies. These sequentially encountered frequencies can be easily and efficiently organized using a natural sampling framework. (Natural sampling structure refers to the non-normalized subset structure that can result from such a tracking and storage system; Kleiter 1994.) Furthermore, this would have included (and still does include) information about item frequencies of significant importance to survival and reproduction: patterns of food distribution, availability of potential mates, of coalition partners and of rivals, frequency of weather events, and frequency of events as indications of time passage. These and many other existing, potentially useful classes of information in the world constituted evolutionary selection pressures to attend to, remember, and think about (De Cruz 2006; Gallistel et al. 2006). Indeed, the entire field of optimal foraging theory in biology is predicated on the assumption that animals have been selected to track quantities of food and space (Stephens and Krebs 1986; see section on comparative evidence). This idea was succinctly expressed by Beran (2008, p. 2):

> Any creature that can tell the difference between a tree with 10 pieces of fruit from another with only six pieces, or between two predators and three on the horizon, has a better chance of surviving and reproducing.

Although our modern, technologically advanced environment contains a great deal of quantitative information expressed in ways other than basic frequencies (e.g., a 60 % chance of rain today), well over 99 % of human evolutionary history did not include such numerical expressions. Over evolutionary history, information about the environment came largely from first-hand experience (even information provided by others would have been restricted to fairly recent events and those people with whom one actually lives). Experientially, single events either happen or they don't—either it will rain today or it will not. So one can observe that it rained on 6 out of the last 10 days with cold winds and dark clouds, and one can have a subjective confidence based on that history, but one cannot *observe* a 60 % subjective confidence. Over evolutionary history, as individuals were able to observe the frequency with which events occur, this information was potentially available to be utilized in decision-making. Thus, if humans have adaptations for inductive reasoning, one might expect them to include procedures designed to take advantage of the existent frequency information in the environment (Hasher and Zacks, Hintzman and Stern 1979, 1978).

What about the apparently non-frequency outputs that people generate, such as a subjective confidence that it will rain *today*? This is not a real dilemma, as it is about the end product of calculations that can easily be based on frequencies up until this point. By analogy, think about opening an image file on your computer. The fact that your computer can produce a picture does not mean that the representational format of that image inside the computer is also a picture. (In fact, it is not.)

4 What Are the Computational Properties of Different Quantitative Representations?

Given that information, even if initially received in the format of frequencies, can be converted into other formats, it is necessary to assess the costs and benefits of different quantitative formats for mental representation. Different numerical representation formats have slightly different mathematical properties, each of which entails certain advantages and disadvantages. What are the computational advantages and disadvantages of frequencies and natural sampling relative to other formats?

Basic frequencies are in many ways the most foundational of numerical formats, both historically and because all other formats are computationally derivable from them. Unlike normalized formats (percentages, probabilities, and ratios), non-normalized frequencies preserve information about reference classes. For instance, *2 out of 3* and *1098 out of 1647* both produce the same normalized numbers (66.7 %, 0.667, 2:5, 2:1), but the first frequency is a much smaller reference class than the later. This can be important in making judgments and decisions, because the size of the reference class provides unambiguous indications of the reliability and stability of the information. Following our above example, the addition of one more number can change *2 out of 3* into *3 out of 4* (drastically changing the nature of the information) whereas the addition of one more number changes *1098 out of 1647* very little (i.e., it is much more reliable and stable information).

Additionally, the storage of information as non-normalized frequencies is important for preserving a high degree of flexibility of categorization and subcategorization. For example, one may have seen 100 people fall sick with an unknown illness, and also observed half a dozen possible diagnostic symptoms (as well as, of course, many non-ill people and their characteristics). Only with the original frequencies is it possible to reconfigure the organizational relationships between illness and symptoms to discover diagnostic patterns. (For example, the 100 sick people could be organized by one symptom, say 25 had a fever, and then reorganized to evaluate a different symptom, say 80 had swollen glands.) Information that has been already normalized to arbitrary reference classes (e.g., 100 for percentages, 1 for single-event probabilities, etc.) would have to be dismantled back into frequencies to be reconfigured in new ways (e.g., given that 25 % of sick people had a fever, the ability to reorganize that information based on a different symptom is lost unless it is reconstituted to the original reference class).

Finally, only with frequencies can new information be easily and automatically incorporated with ongoing experiences. Once again, normalized numbers obscure the size of the reference class and make it impossible to tell how much (or how little) a single new observation will change the nature of the data. Frequencies are simply easier to work with because they conserve information about the base rates (Kleiter 1994).

There most definitely are certain situations in which naturally sampled frequencies appear to be unable accomplish particular tasks. Two such situations that are often raised are (a) judgments about novel, one-time events, and (b) determining if observations are statistically independent of one another (Over and Green 2001).

With respect to one-time events, one possibility is that people look to similar, though not identical, events for guidance in these situations;[1] that puts these cases back into the realm in which natural frequencies can be utilized. An alternative possibility is that, because singular events in one person's lifetime may have occurred repeatedly across the evolutionary history of a species, the mind might contain evolved behavioral predispositions regarding that type of event (assuming the event had inclusive fitness consequences). This would not be based on natural frequencies, but it also is clearly not free of evolutionary or ecologically rational processes either. Finally, it is possible that—as some people do find when faced with a novel, one-time event—one simply does not have clear or consistent guidance about what to do. The second situation (of determining statistical independence of observations) is, indeed, easier with normalized numbers. These normalized numbers, however, are quite obtainable as derived from frequencies. There is no *a priori* reason why these insights cannot be derived from the frequency information contained within naturally sampled frequencies, following an algorithmic transformation. There is, however, an insurmountable problem if only the percentage information is available. Because information about the reference class size is gone, one cannot say whether or not the information (e.g., 80 %) is based on 800 out of 1000 or based on 4 out of 5.

5 What About Other Animals?

Many of the initial considerations of an ecological rationality approach apply not only to humans but to other animals as well (e.g., the fit between the structure of the natural environment and the structure of the [non-human] mind; time and computational constraints of a situation; the inherent statistical properties of the natural world within which the choice is being made). It is therefore important to look at phylogenetic evidence for how quantities are represented in non-human animals. This includes information from comparative psychology, primatology, physical anthropology, and paleontology.

A number of evolutionary biologists have investigated behaviors that involve judgment under uncertainty in non-human animals in order to test various mathematical models from optimal foraging theory (e.g., Stephens and Krebs 1986). Their

[1] A related option is that people may be able to construct frequentist (or quasi-frequentist?) probabilities by counting hypothetical possibilities in the sample space: using theoretical frequencies, rather than actual frequencies. In other words, imagining the possible outcomes in a way that still respects constraints of the real world (see Shepard 1984 for a similar idea within the field of perception). For example, a judgment about how likely it is to roll 10 or more for the sum of two 4 sided dice and one 8 sided die is probably a novel one-time event, but one can figure it without too much effort. (Either by using subjectivist methods of expected value, or by counting in the hypothetical frequencies within the relevant sample space.) The expected value solution would be that I expect to roll an average of 9.5 for the sum of the dice, so rolling a 10 or more should be 50 %. Investigating how people, in fact, deal with situations such as this is important but not adequately studied at this time. The most immediate point for our purposes here is that these situations posed by Over and Green (2001) are far from conclusive.

experiments have demonstrated that a wide array of animals—even ones with truly minuscule nervous systems, such as bumblebees—make judgments under uncertainty during foraging that manifest exactly the kind of well-calibrated statistical induction that humans have seemingly struggled with on paper-and-pencil tasks (e.g., Gallistel 1990; Real 1991; Real and Caraco 1986; Staddon 1988). How can this be?

At a basic level, a wide array of non-human animals (including salamanders, rats, various types of birds, dolphins, monkeys, and apes; see Beran 2008 for an overview) are sensitive to the quantitative properties of various natural events. These are, to be sure, experienced frequencies of objects, events, and locations, rather than explicit numerical notations (most animals do not understand written numbers). But even in the realm of explicit numerical notation, non-human primates have been shown to be capable of simple mathematic operations (e.g., matching, addition, and subtraction of small quantities) using abstract stimuli such as the number of dots on a computer screen (Beran and Rumbaugh 2001; Boysen and Berntson 1989; Brannon and Terrace 2000, 2002; Cantlon and Brannon 2007; Inoue and Matsuzawa 2007; Matsuzawa 1985; Tomonaga and Matsuzawa 2002).

What is one to make of the contradiction between these impressive non-human animal performances and the often seemingly error-prone human judgment and decision making abilities? A key to resolving this apparent paradox is to realize that these other animals (including even bumblebees) made great judgments under uncertainty because they were tested under conditions that were ecologically valid for their species (Tooby and Cosmides 1992). The ecological rationality approach simply claims that the same should hold true for humans.

6 What Is the Cognitive Psychology of Quantitative Information?

There are several areas of research within Psychology that provide relevant evidence about how people encode, represent, and use quantitative information. This is also, perhaps not coincidentally, the area in which the disagreements between ecological rationality and heuristics and biases views get played out most often and most assiduously. This section is divided into parts, discussing the research on how well people can track the occurrence of frequencies, how people evaluate different numerical formats, how people perform on complex statistical tasks when given information in different numerical formats, and how adding different pictorial aids can influence statistical reasoning.

6.1 Frequency Tracking

Although frequency judgments in naturalistic studies of autobiographical memory have been inaccurate in many studies, laboratory studies that more carefully control for the multitude of potentially confounding factors endemic to naturalistic studies have found remarkably accurate frequency judgments (for a review, see Hasher and Zacks 2002). Even studies using classic tasks that are supposed to demonstrate the fallibility of frequency estimation abilities (e.g., by people having to resort to the availability heuristic; Tversky and Kahneman 1973), have been found to nevertheless maintain accurate relative frequencies (Sedlmeier et al. 1998). This work has also led to a more subtle point that the ecological rationality perspective does not predict perfect or optimal frequency tracking in terms of explicit numbers, but rather that frequency tracking should be expected to *satisfy*: to be as accurate as necessary under normal ecological circumstances (i.e., realistic item and computational limits) so as to produce good judgments and decisions (see Simon 1956).

6.2 Evaluating Different Numerical Formats

Another way to evaluate and understand quantitative information is to directly ask them. Brase (2002a, 2002b, 2002c) found that when people are asked to evaluate the clarity and understandability of different statistical statements, frequencies (both simple frequencies and relative frequency percentages) are seen as clearer and easier to understand than single-event probabilities. This research also found that simple statements of large-scale frequencies (e.g., millions or billions) lead to systematic distortions in the persuasive impact on both attitudes and potential behaviors, both for very low base rates (e.g., expressing a 1 % base rate in terms of absolute numbers of the world populations, which leads to greater relative impact than other formats) and for very high base rates (e.g., expressing a change from 6 billion to 7 billion, which leads to less relative impact than other formats).

Wang (1996a, 1996b); Wang and Johnston (1995); Wang et al. (2001) similarly found that using large-scale frequency contexts led to systematic differences in behavior. In this case, when people were asked to make decisions about a small-group or family-sized populations (fewer than 100 people) the traditionally observed framing effects of decision malleability (i.e., shifting towards riskier outcomes when framed as gains) disappeared (Tversky and Kahneman 1981; Wang 1996a, 1996b; Wang and Johnston 1995). The explanation for this pattern of sudden consistency stemmed from the ecological rationality insight that these smaller population sizes are on scales of magnitude with which humans have directly and recurrently dealt over their personal history.

6.3 Statistical Reasoning with Different Numerical Formats

A classic area of research on statistical reasoning is that of Bayesian inference (determining the posterior probability of an event, given some new information to be combined with an initial base rate; for example, how likely a patient has a disease given a positive test result), with the classical finding that people very rarely perform well (i.e., reach the correct Bayesian inference; e.g., Casscells et al. 1978). This classical finding, however, was based on tasks in which the information was presented in probabilistic formats (most often, percentages and single event probabilities).

When people are given information as naturally sampled frequencies, however, they are more likely to generate normatively correct statistical inferences (e.g., Gigerenzer and Hoffrage 1995). Research has demonstrated that naturally sampled frequencies (often shortened to just "natural frequencies") can be used to improve the communication and understanding of medical statistics by physicians and patients, judges and jurors, and even children (Hoffrage et al. 2005, 2000; Zhu and Gigerenzer 2006).

Disputes still exist about the theoretical implications of these effects for our human cognitive architecture, and specifically whether or not it implies that the human mind is inherently predisposed to process natural frequencies (Brase 2008). Distinguishing between these two theoretical interpretations is difficult because the statistical reasoning tasks traditionally used in this research involve exactly the sorts of computations that get simpler with the use of natural frequencies (Gigerenzer and Hoffrage 1995).

The claims that certain formats—most often percentages or chances—are just as clear and easy to understand as frequencies has been a consistent critique of the ecological rationality view. It is also, however, a critique that rests on very specific and tenuous arguments. Specifically, the claim that chances (e.g., a person has 20 chances out of 100) are probabilities and are therefore importantly different from frequencies (e.g., 20 people out of 100) was promulgated by Girotto and Gonzalez (2001) and argued against by ecological rationality proponents (Brase 2002b; Hoffrage et al. 2002). The crux of the claim is that such chances refer to a single event (e.g., a person) and therefore must be a single-event probability. The counter-argument is that such statements are really about the multiple chances (be they real or hypothetical), and therefore these chances can be just as validly defined as frequencies. Recent research to clarify this issue (Brase 2008) found that research participants—when asked—did report differing interpretations of the "chances" numbers in Bayesian reasoning tasks (some interpreted them as frequencies; others as single-event probabilities). These participants, however, performed better when given equivalent frequencies as compared to isomorphic tasks that involved chances. Furthermore, the participants who were given tasks that used the "chances" wording performed better if they had adopted a frequentist interpretation of those chances. Thus, the assertion that chances are probabilities, and not frequencies, is something that not only many researchers reject but many research participants also reject. The assertion that performance with chances is equivalent to performance with natural frequencies simply does not survive serious scrutiny.

The issue of how people consider percentages is, like the issue with chances, confused by the ambiguous nature of the format. Generally, percentages can be more formally considered normalized relative frequencies, that is, they are an expression of frequency out of a normalized quantity (100). Thus, 37 % means "37 out of 100". This definition of percentages places it within the family of frequencies, but not as naturally sampled information. Any claim that percentages are better (or worse) than frequencies has to be implicitly considering percentages as being defined in a way slightly different from that definition (e.g., considering percentages as expressions of subjective confidence). There is a second problem with the claim that percentages (of some conception) are just as clear and easy to understand as frequencies; it is based on a faulty reading of the research. Barbey and Sloman (2007, p. 249) argued for the status of percentages with perhaps the most forceful argument:

> . . . single numerical statements. . . have a natural sampling structure, and, therefore, we refer to Brase's "simple frequencies" as natural frequencies in the following discussion. Percentages express single-event probabilities in that they are normalized to an arbitrary reference class (e.g., 100) and can refer to the likelihood of a single-event (Brase 2002a; Gigerenzer and Hoffrage 1995). We therefore examine whether natural frequencies are understood more easily and have a greater impact on judgment than percentages.

As pointed out by several commentators (e.g., Brase 2007) this interpretation is only possible due to a remarkable combination of incorrect reinterpretations (for instance, re-labeling simple frequencies as natural frequencies), incomplete definitions (of percentages, which *can* refer to single events, but in the more usual case refer to relative frequencies), and (for the particular case of the Brase 2002a results) the omission of experimental results that would have invalidated the claim.

6.4 Statistical Reasoning with Natural Frequencies and with Pictures

The use of pictorial representations to aid in judgment and reasoning tasks has been an area of some recent interest. The ecological rationality explanation for performance improvements due to supplementary pictures is that pictures further push participants "to represent the information in the problem as numbers of discrete, countable individuals" (Cosmides and Tooby 1996, p. 33). Other theorists, from a more heuristics and biases position, have explained the pictorial representation boost as a result of the pictures making "nested set relations transparent" (Barbey and Sloman 2007, p. 248; Sloman et al. 2003). Barbey and Sloman also point out that "abstracted, pictorial representation (e.g., Euler circles and tree diagrams) have been shown to improve performance across different deductive inference tasks such as categorical syllogisms and the THOG task, as well as categorical inductive reasoning tasks" (p. 251), which they count as a finding in favor of dual processes theories even though it could legitimately be explained from either perspective.

Recent research has pitted these two interpretations of what aspects of pictorial representations are producing facilitation in judgment under uncertainty tasks. By comparing different types of pictorial representations (Euler circles, rows of icons, and circles with dots in them), Brase (2009) was able to establish that the best and most consistent facilitation due to pictorial representations was from the use of discrete, countable icons. This lends greater weight to the ecological rationality interpretation of pictorial representation facilitation, and thus the privileged representation status of frequencies (and frequency-like pictorial presentations).

6.5 What About Infants and Children?

If, as the ecological rationality position proposes, people have an inherent tendency to encode, store, and use quantitative information in the form of frequencies then this should be apparent even in the earliest stages of cognitive development. Evidence actually exists of this specific progression across development (in typical environments, in the absences of developmental disorders).

Developmental work has demonstrated surprising quantitative abilities in infants, so long as the information is presented in ecologically realistic, frequentist ways. (It is interesting to note some clear similarities between the infant cognition methodologies and the comparative psychology methodologies described above.) Using methods such as the habituation paradigm (e.g., Xu and Spelke 2000) and visually presented objects and events, it is possible to "ask" infants questions such as, "If there were two objects, and one is removed, should there be anything left?" (i.e., $2 - 1 = ?$). Infants within these paradigms are remarkably good at doing simple addition, subtraction, and comparisons (Gallistel and Gelman 1992; Gilmore et al. 2007; McCrink and Wynn 2004; Van Marle and Wynn 2011; Wynn 1998a, 1998b; Xu and Spelke 2000). Infants can also perform matching to samples (Xu and Garcia 2008).

Such clear, consistent, and dramatic results have led many researchers to propose that there is a core "number sense" that is part of the normally developing human endowment (Feigenson et al. 2004; Lipton and Spelke 2003; Spelke and Kinzler 2007; for dissenting views, see Mix et al. 2002; Mix and Sandhofer 2007). As children get older, this basic number sense appears to become elaborated upon through education (Sophian 2000). When we reach K-6 level children, as discussed above, use of natural frequencies in Bayesian reasoning tasks becomes relevant. When information is presented in the form of natural frequencies, even 5th and 6th graders can successfully solve simple Bayesian reasoning problems (Zhu and Gigerenzer 2006).

6.6 Are Quantitative Abilities Consistent Across Cultures?

Cross-cultural lines of evidence for a trait hinge on the twin issues of universality and variation. The psychological and physiological bases of a human trait should be reliably developing in virtually all cultures, but at the same time there can be variations in the expression of traits (i.e., ecology-dependent variability and facultative/conditional adaptations). For example, it is cross-culturally universal that females are capable of bearing children, but there is variation in cultural circumstances that affect rates of actual per-individual children borne.

Deliberate cross-cultural research to evaluate ecological rationality explanations for statistical reasoning has not been done. Much of the research on frequency representations in statistical reasoning has spanned several countries and cultures (United States, Germany, England, etc.), but these have been relatively homogeneous in that they are all technologically advanced, Western cultures. Research on Asian/Western differences in general mathematic skills has focused on the acquisition of academic mathematical abilities more generally in students. It remains to be seen what variations exist between Asian and Western cultures when ecologically rational tasks are systematically used. The ecological rationality perspective predicts that a number of the already established differences between these cultures may be reduced when using such tasks.

Although general cross-cultural research can be tremendously helpful, more targeted research on hunter–gatherer societies is particularly important. Hunter–gatherer societies provide an approximation of ancestral human environments, both including information and environments that are no longer common in industrialized societies and excluding information and environments that have only recently become common. It is therefore important to specifically look at hunter–gatherer societies (e.g., via cultural anthropology, human ethology, and human behavioral ecology) as a potential line of evidence.

Comparisons of numerical representation abilities across technologically advanced cultures and hunter–gatherer (or slash and burn horticulturalist) societies are complicated by the radically different educational systems across these cultures. In technologically advanced cultures, there is a vast array of *explicit* quantitative information—numbers—in the surrounding environment. Numbers are found on telephones, computers, digital clocks, timers, product labels, newspapers, magazines, television, radio, the Internet, and so forth. These explicit numerical representations are largely or entirely absent in hunter–gatherer cultures. Nevertheless, and perhaps remarkably so, mental representations of quantity are ubiquitous across cultures of all types. Brown (1991) summarizes the general state of numerical representation across all cultures thus:

> People may have a very elementary system of numbers and yet have a full-blown ability to count (which allows them very quickly to adopt complex number systems when they become available and prove useful). Theoretically, Hall (1975) argues that numbers and counting could be absent among a given people, particularly if they had no need to count. And yet the ability to count is universal as an innate (and presumably specific) capacity of the human mind. (p. 46)

One implication of this situation is that the traditional paper-and-pencil statistical reasoning tasks are unsuitable for research with most hunter–gatherer populations. The range and scale of explicit numbers, however, pales in comparison with the amount of quantitative information that is *implicit* in the world around us: All the objects, events, and locations in the world that can be perceived as multiple, discrete items are potential contributors to implicit quantitative information. If one can track either how often something occurs or how many there are (consciously or subconsciously), it is a source of quantitative information.

All of this quantitative information that exists around us is potential fodder for individuals to use in order to understand and interact with the environment. From lab animals that experience different schedules of reinforcement in operant conditioning (tracking the frequency of rewards along with the frequency of the operant behavior), to wild animals that adjust foraging behaviors in response to the frequency with which food is found in different locations, to people who notice the frequency of cars that go by as they walk along the street (perhaps considering when to cross), individuals track and use quantitative information for successfully navigating the world—even in the absence of explicit numbers.

There does appear to be one culture in which there seems to be a profound lack of any explicit, or even implicit, number system. Members of the Pirahã tribe—a group of South American hunter–gatherers—seem to have no concept of numbering and counting, even when explicitly tested for the existence of such an ability (Frank et al. 2008; Gordon 2004). The Pirahã, furthermore, show profound difficulties in tracking and maintaining visual representations of quantities above about 4 items. This suggests that an ability to track and utilize frequencies of objects, events, and locations may not actually be a universal human ability, under evolutionarily representative circumstances.

This case of the Pirahã tribe presents an intriguing situation, and further work with the Pirahã is certainly warranted although this is difficult given that there are only about 300 members of this tribe. What is particularly perplexing is the long list of abilities that they apparently fail to exhibit: They reportedly have no distinct words for colors, no written language, don't sleep for more than two hours at a time, communicate almost as much by singing, whistling and humming as by normal speech, frequently change their names (because they believe spirits regularly take them over and intrinsically change who they are), do not believe that outsiders understand their language even after they have just carried on conversations with them, have no creation myths, tell no fictional stories, and have no art. One can easily get the impression that there is either a more general and pervasive cognitive issue involved with the Pirahã, or that there may be something unusual going on in the social or methodological details of researcher/Pirahã interactions. One of the foremost experts on the Pirahã has ascribed a large number of these characteristics to a feature of Pirahã culture that "constrains communication to nonabstract subjects which fall within the immediate experience of interlocutors" (Everett 2005, p. 621). What is clear is that more, and better controlled studies are needed on the quantitative representational abilities (including implicit representations) of non-industrialized populations (Casasanto 2005; Gordon 2005; see also Dehaene et al. 2006 for related research on a very similar group of Amazonian hunter–gatherers).

7 Is There a Neuroscience of Quantitative Representation?

The claim that the mind is designed to more effectively take in, process, and understand frequency representations of quantitative information directly implies that there should be specific neuroanatomical structures that are dedicated to performing this function. Although the judgment under uncertainty literature had not generally referenced the neuroimaging literature, there is a wealth of support for this claim.

Both functional neuroimaging studies and studies of patients with selective brain damage indicate that the intraparietal sulcus (IPS) and prefrontal cortex (PFC) are key areas related to the processing and tracking of quantitative information (see Butterworth 1999; Dehaene 1997 for reviews). Furthermore, there is evidence that numerical processing ability, in those locations, develops early in the lifespan (at least by three months of age; Izard et al. 2008). Other work has demonstrated that analogous brain regions are responsible for numerical information processing in non-human animals (e.g., Diester and Nieder 2007).

There is still debate as to the particular roles of these different areas, but at this point it appears that the PFC is more involved in the mapping of symbolic representations with numerical concepts, whereas the IPS is more directly involved in the tracking of numerical information in earlier stages. With increasing age and numerical proficiency, the understanding of numerical information (whether symbolic numbers, numbers words, or analog stimuli) appears to rely less on the PFC and more directly on the parietal cortex (Ansari et al. 2005; Cantlon et al. 2006; Cohen Kadosh et al. 2007; Piazza et al. 2007; Rivera et al. 2005).

There is also debate at this point about the specific nature of these frequency tracking abilities; whether they are composed of one basic number system or two basic systems (one for small sets and another for large sets; Beran 2007; Cantlon and Brannon 2006; Feigenson et al. 2002; Feigenson et al. 2004; Xu 2003). What is notable for present purposes is that the idea of neurological underpinnings for tracking and storing numerical quantities, established as a fact, is now the background assumption for this debate (Ansari 2008; Jacob and Nieder 2009).

8 A Summary of the Evidence

The preceding sections outlined affirmative evidence for the question of if frequentist interpretations of probability privileged in the human mind, as suggested by an ecological rationality perspective. Specifically, this privileged frequentist thesis derives quite strong theoretical support from evolutionary theory, information theory, cognitive psychology, developmental psychology, cross-cultural research, and neuroscience (see Fig. 1).

The converging lines of evidence shown in the figure helps to identify and organize various bodies of evidence applied to this topic, connecting initially disparate phenomena into integrated scientific understanding, and allowing concise comparisons between the views. These are useful accomplishments for researchers, for scientific communication, and for teaching. (See Schmitt and Pilcher's 2004 model for a similar approach to organizing multiple converging lines of evidence.) In summary, when assessed in terms of multiple, independent, and converging lines of evidence, the privileged frequentist hypothesis appears to be a very good scientific description of reality.

Fig. 1 Summary of the ecological rationality view of human numerical representation, using the convergence framework

9 What Does This All Mean for Mathematics Education?

What are the educational implications of having a mind built to preferentially think about frequencies, rather than other numerical formats? First, and most obviously, it implies that other numerical formats should often be problematically interpreted through a frequentist lens. The first part of this section reviews some of the evidence that these phenomena actually occur. We then look at how these phenomena can conceptualized and applied in the service of better mathematics education, and finally, we conclude with a summary of an ecologically rational pedagogical model of mathematics education.

9.1 Pervasive Frequency Interpretations

First of all, learning of frequencies is developmentally primary (i.e., it happens first), it is relatively unproblematic, and it is therefore the foundation for other numerical concepts. Of course, children can also learn other numerical representations such as percentages, fractions, decimals, and even single-event probabilities. In contrast to basic frequencies, though, these representations require more effort to be learned (Gallistel and Gelman 1992), and in the process of learning these other formats there are persistent problems with children trying—incorrectly—to understand these other representations as frequencies. Research indicates that students need to be competent in the four basic operations of whole numbers and have a solid understanding of measurement for them to be prepared to learn rational numbers (Behr and Post 1988). Rational numbers are the first such numbers experienced by students that are not based on a counting algorithm. This shift is a challenge for many students. Initially, whole number knowledge inhibits learning rational numbers because children over generalize their counting principles and because whole numbers have a "next" number whereas rational numbers do not (Behr et al. 1984; Carraher 1996; Gelman and Meck 1992).

For example, there is a long-documented tendency for children to develop "buggy algorithms" in their understanding of fractions and decimals (Brown and Burton 1978; Gould 2005; Lachance and Confrey 2001; Resnick et al. 1989). A couple of the most common example of these "bugs" are:

(a) The rational number addition bug (a.k.a. the freshman bug), which involves adding the numerators and the denominators of two fractions when trying to combine them (e.g., $1/3 + 1/2 = 2/5$; Silver 1986).
(b) The decimal version of the rational number addition bug is sometimes called "Benny's bug" (Erlwanger 1973). An example of this is when a student encounters a decimal math problems such as "$0.2 + 2.0 =$?" and responds with "0.4". What they have done is add together the integers as whole numbers and then placed them after a decimal (presumably because they recognize that this is a decimal problem).

What these bugs have in common, other than their frequency of occurrence, is that they reveal a tendency to treat numbers—even when expressed as fractions and decimals—as frequency counts. If "$1/3 + 1/2 =$" were referring to actual counts (e.g., 1 out of 3 apples are green, another 1 out of 2 apples are also green, and we put all the apples together in a pile, the answer is indeed 2 out of 5 green apples). There are very specific properties of both fractions and decimals that preclude that interpretation (Brase 2002c), but those properties do not seem to be as intuitively obvious as the frequentist interpretation and they are too often not clearly taught as the explicit properties of these numerical formats. For example, Kieren (1976) identified seven interpretations of fractions that formed the foundation for solid rational number knowledge. These were ultimately condensed to five distinct, yet interconnected, interpretations of fractions: Part/whole, measure, operator, quotient, and ratio (Kieren 1988; Lamon 2001). Clements and Del Campo (1990, p. 186) note that:

> At various times of their schooling, children are told that the fraction $1/3$, for instance, is concerned with each and all of the following: (a) sharing a continuous quantity between three people, (b) sharing 12 (say) discrete objects between three people, (c) dividing the number 1 by the number 3, (d) a ratio of quantities, (e) a 1 for 3 replacement operator, (f) a rational number equal to $2/6$, $3/9$, and so forth, and (g) a decimal fraction of 0.333.

Decimals are usually taught after fractions in mathematics curricula, and thereafter becomes the predominant format (Hiebert and Wearne 1986; Hiebert et al. 1991). Yet the occurrence of "Benny's bug" indicates that even students who have learned fractions are still predisposed to frequentist interpretations of numbers. In fact, the learning of decimal fractions seems to lead to a proliferation of buggy conceptions (Brase 2002c; Moloney and Stacey 1997; Nesher and Peled 1986; Resnick et al. 1989; Sackur-Grisvard and Leonard 1985; Steinle 2004), all of which can be understood as attempts to apply a frequentist view to decimals:

(a) Adopting the rule that "more digits means bigger" (e.g., 0.1234 is larger than 0.32) generally occurs because students are using a judgment method that was successful for whole numbers.
(b) Adopting the rule that "more digits means smaller" can be a reaction to learning that the first rule is wrong (e.g., 0.4321 is smaller than 0.23), or it can result from the use of place value names to decide on the size of decimals. That is, they recognize that one tenth is larger than one hundredth and reason that any number of tenths is going to be larger than any number of hundredths. They can also respond with this rule if they are confusing notation for decimals with that for negative numbers. When comparing 0.3 and 0.4, they will reason that -3 is greater than -4.
(c) Adopting a rule that attaching zeros to the right of a decimal number increases the size of that number (e.g., $0.8 < 0.80 < 0.800$) by treating the decimal part as a whole number.
(d) Adopting a rule of ignoring zeros on the left (e.g., $0.8 = 0.08 = 0.008$) in a way similar to the "more digits means bigger" rule with a variation. Just as the number 8 is not changed by placing a zero in front of it (08), the zeros after the decimal point are also disregarded making 0.8 the same as 0.08 just as 8 is the same as 08.
(e) Adopting a rule of ignoring decimals, thus lining up digits on the right rather than lining up the decimal points (e.g., Benny's bug: $0.2 + 4 = 0.6$, and variants: $0.07 + 0.4 = 0.11$, $6 \times 0.4 = 24$, and $42 \div 0.6 = 7$).

As Hiebert and Wearne note (1986, pp. 204–205):

> Extending concepts of whole numbers into referents that are appropriate for decimal fraction symbols is a delicate process. Students must recognize the features of whole numbers that are similar to decimal fractions and those that are unique to whole numbers.... The evidence suggests that many students have trouble selecting the features of whole number that can be generalized.... Most errors can be accounted for by assuming the student ignores the decimal point and treat the numbers as whole numbers.

The transition from basic frequencies to normalized frequencies is a matter of complex changes in perspective. Early stages in the transition involve tracking two frequencies that are repeatedly added. So that "2 out of 3" is equivalent to "4 out of 6" because a repetition of three and a repetition of two generate an equivalent value. Similarly 37 % can be interpreted as 37 out of 100, or 370 out of 1000. The "*m* parts out *n* parts" that describes additive thinking leads students to think that the numerator is a part of the denominator. Fractions such as $3/2$ are thus difficult to interpret because this involves thinking of 3 as part of 2 (Thompson and Saldanha 2004). Students must learn to interpret fractions as a single number rather than thinking of them as two numbers (Kerslake 1986). As a student's schooling progresses, and fractions begin to be treated as numbers themselves, a key transition is that students move away from "so many out of so many" and begin to think of fractions multiplicatively rather than additively (Thompson and Saldanha 2004; Vergnaud 1994).

This process begins with the coordination of partitioning, multiplication, and measurement (interpreting "A is m/n times as large as B" as meaning that A is m times as large as an nth part of B), a perspective now a critical part of the Common Core State Standards in Mathematics. It ends when a student's understanding of fractions is fully integrated with conceptions of proportion and the real number system so that a student can imagine a fraction as a smooth proportion: a relationship between two quantities that change continuously though all real values, but are coordinated in such a way that one is always a constant number of times as large as the other (Castillo-Garsow 2010; Thompson and Thompson 1992) just as an integer is the result of all possible subtractions that generate that integer (Thompson 1994).

In the case of Ann (Thompson and Thompson 1994, 1996), this distinction between adding on (additive reasoning), and cutting up (multiplicative reasoning) can be seen to have a direct impact on the ability of a student to solve problems of distance, time, and speed. Ann thought of speed as a repeated distance: a speed of 20 feet per second meant repeatedly adding 20 feet every second. Ann was able to solve unknown time problems by counting the number of times the 20 feet speed–length fit into a total travel distance of 100 feet. However, the problems of finding an unknown speed from a distance and time had her stumped because she did not have a speed–length to count. It was not until a researcher intervened to build a meaning of speed based on the cutting up of accumulated distance and accumulated time (Cutting up 7 seconds into 7 pieces and 100 feet into 7 pieces to arrive at a speed in feet per second) that Ann was able to solve these problems. We can see in this example a case of the student making a transition from additive reasoning to multiplicative reasoning and building a new understanding of rate and fractional equivalence. She has come to imagine that 100 feet in 7 seconds is the same as 100/7 feet in 1 second, and in doing so she has built equivalence between basic natural frequencies.

9.2 Making Sense of the Bugs

The key insight from an ecological rationality approach is that these problematic phenomena have a clear and comprehensible explanation. Whereas earlier explanations have tended to view these buggy procedures as developing spontaneously due to overgeneralizations of earlier learned mathematics (whole numbers) to new topics (decimals/fractions; e.g., Resnick et al. 1989), the ecological rationality perspective explains these phenomena as a part of how the human mind is designed to work. This also, of course, helps to explain the frequency and pervasiveness of these buggy procedures across classrooms, topics, and generations of students. What has been described as a "whole number bias" (e.g., Ni 2005) is not merely an annoyance and a buggy bias to be stomped out; it is a design feature of the human mind. As such, pedagogical techniques will be more fruitful (and less frustrating) is they take those design features into account and work with them in the teaching of mathematics.

A number of the considerations and lines of evidence in the preceding sections have led to theoretical models for the development of mathematics. For instance, Geary (1995) proposes conceptualizing mathematics into two fundamental types: biologically primary mathematical abilities and secondary abilities which develop on top of those primary abilities. This has also been extended to, for instance, the consideration of mathematical disabilities (e.g., dyscalculia; Geary 2007; Geary et al. 1999). Given the history of a heuristics and biases perspective dominating ideas in mathematics education, it is perhaps not too surprising that this more ecologically rational model has met with some resistance (e.g., Glassman 1996).

The Common Core State Standards (CCSS) places a lot of attention on fractions while, some would say, snubbing decimals. In grades 4 and 5, the symbolic notation of decimals is viewed as a notation for a particular type of fraction (specifically with base-ten denominators). While this marks a clear attempt to avoid the misconception that decimals are something totally different, a concerted effort to help students see this will be necessary. Assisting students as they come to terms with the complexity of decimals is often undervalued and underemphasized. Although students struggle with rational numbers, in general, decimals are one of the greatest challenges. Results from the Third International Mathematics and Science Study (TIMSS) indicate that students do not perform as well on questions involving decimals compared to those involving fractions (Glasgow et al. 2000). The decimal representation is compounded in complexity by the merging of whole number knowledge and common fractions with very specific kinds of units. In addition, decimals can be viewed as both continuous and discrete.

The impact of whole number (i.e., frequency) thinking on students' development of decimal number knowledge may also relate to the nature of the knowledge. Students' prior knowledge may be predominately procedural, which may account for misconceptions applied to decimal numbers (Hiebert and Wearne 1985). Research suggests the same is true for the conceptual understanding for fractional operations. Students have a procedural knowledge of operations rather than an understanding of fundamental concepts that form the basis for fractions (Mack 1990). Often students inappropriately apply rules which can result in the right answer for the wrong reason. This reinforces the inappropriate use of the rule, and the error rules persist and procedural flaws are not corrected (Hiebert and Wearne 1985; Steinle and Stacey 2004b). Students who are instructed using only procedural methods tend to regress in performance over time (Woodward et al. 1997; Steinle and Stacey 2004a). In addition, knowledge developed in one representation (e.g., decimals) does not necessarily transfer to all representations of rational numbers (Vamvakoussi and Vosniadou 2004).

Kieren (1980) suggests that partitioning is fundamental to the meaningful construction of rational number just as counting is the basis to understanding the construction of whole number. In fact, the construction of initial fraction concepts hinges on the coordination of counting and partitioning schemes (Mack 1993). The notion that partitioning results in a quantity that is represented by a new number is fundamental to rational numbers. For many children, coordinating their ideas to reach this level takes time and experience. For example, Mack (1995) found that third and fourth graders use their understanding of partitioning to connect operations on fractions to their prior knowledge of whole numbers. This enables them to solve addition and subtraction problems with a common denominator by separating the numerator and denominator and thinking only about the number of pieces combined or removed. Although this process assists them with the operation, it limits their conceptions about fractions by treating them as whole numbers. The same is true with decimal numbers where children separate parts from the number and treat them as a whole number. Instruction on decimals as an extension of the whole number system often occurs without an adequate understanding of place-value concepts that enable them to work with whole numbers (Fuson 1990). As a result, students often attempt to write too many digits into a column. Extending the place value structure to include digits to the right of the decimal point presents additional challenges. Without understanding the value of the columns, some students think the further away from the decimal point, the larger the value of the digits (e.g., 0.350 is larger than 0.41 because 350 is larger than 41).

Applying whole number properties inappropriately also occurs because students fail to understand the symbolic representation of fractions, which is often prematurely introduced. Research indicates that students base their informal knowledge of fractions on partitioning units and then treat the parts as whole numbers (Ball 1993; D'Ambrosio and Mewborn 1994; Mack 1990, 1995; Streefland 1991). Without a meaningful understanding for fraction symbols, students acquire misconceptions from attempting to apply rules and operations for whole numbers to fractions. Mack (1995) indicates that, with time and direct effort, students can separate whole number from rational number constructs and develop a meaningful understanding of how fractions and decimals are represented symbolically. Instruction must begin with a focus on student transition from additive to multiplicative reasoning and move students from seeing fractions and decimals as two number to seeing them as a single value (Kent et al. 2002; Sowder et al. 1998). Students can then build relationships among fractions, ratios and proportions (Sowder et al., 1998). Fraction concepts and the relationship among various forms of rational numbers can be explained through the use of a combination of representations such as verbal statements, images/pictures, concrete materials, and real world examples, and finally written symbols. Students have to develop appropriate images, actions, and language to set the stage for formal work with rational numbers (Kieren 1988; Lesh et al. 1983).

The entrenched tendency of students to over-apply frequentist whole number knowledge is has been recognized in education research and often interpreted as being due to primitive schema that are deeply embedded and difficult to modify (McNeil and Alibali 2005). On this interpretation, it is the persistence of the over-applications and resulting misconceptions, the resistance to change despite instruction, that presents a problem (Harel and Sowder 2005; Steinle and Stacey 2004a; Zazkis and Chernoff 2008). The experiences that contradict previous knowledge and create cognitive conflict are often avoided when students adopt "coping strategies" such as annexing zeroes to equalize the length of decimal numbers in comparison situations or blindly using left to right digit comparison procedures. Students often adopt these procedures temporarily but then later regress to behaviors consistent with their prior schema (Siegler 2000; Steinle 2004).

Another example is descriptions of the "stickiness" of additive reasoning (rather than considering that this may be an inherent predisposition towards frequentist interpretations). Because students are familiar with the concept of addition, and they learn to rely on it early and often, this is assumed to be the reason why it is often very resistant to change. Indeed, additive reasoning is used in a qualitative, intuitive way, not just by seven-year-olds, but by students from the ages of 11 through 16 who have been taught something about proportions (Hart 1981). Furthermore, research shows that children do not "grow out" of erroneous addition methods (Thornton and Fuller 1981) and that this additive error may actually cause a delay in the development of multiplicative thinking (Markovits and Hershokowitz 1997) which can disrupt work with rational number and ultimately with more advanced topics such as slope and probability. For example, students inappropriately apply additive thinking when judging the equivalency of two fractions with different numerators and denominators (Behr et al. 1984).

Zazkis and Chernoff (2008) suggest that students who wrestle with counterexamples that enable them to personally experience potential cognitive conflict, are more likely to respond with new learning than those where expert opinion is simply staged for them. Contributing to the complexity of rational numbers are the concepts of equivalence and that rational numbers can be represented in various forms and can look very different but mean the same thing.

We would be remiss if we didn't address percent as an equivalent form of a rational number. Percent is a particular way to quantify multiplicative relationships. According to Parker and Leinhardt (1995), it "is a comparative index number, an intensive quantity, a fraction or ratio, a statistic or a function" (p. 444). Throughout all of these interpretations, it is "an alternative language used to describe a proportional relationship" (Parker and Leinhardt 1995, p. 445). Together these interpretations create the full concept of percent and are essential understandings in order to solve a wide variety of problems involving percent (Risacher 1992). As with the other interpretations of rational number, students often use incorrect rules and procedures related to percentages (Gay and Aichele 1997). This may be a direct consequence of students studying from a curriculum that emphasizes rules and procedures (Gay and Aichele 1997; Hiebert 1984; Rittle-Johnson et al. 2001). When they are not sure what to do, students will revert to rules and procedures from concepts that are more familiar, more intuitive, and/or resistant to change, such as whole number (Risacher 1992).

Common errors working with percent have been identified in research (Parker and Leinhardt 1995; Risacher 1992). First, students tend to ignore the percent sign and treat the percent as a whole number. Second, they follow what is referred to as the "numerator rule" where they exchange the percent sign on the right with a decimal on the left. Third, they implement the "times table", also known as a "random algorithm" (Parker and Leinhardt 1995; Payne and Allinger 1984; Risacher 1992). In addition to problems that arise from overgeneralizing whole number rules and procedures, misconceptions also result from students' limited instruction of percent being part of a whole. Researchers regard the interpretation of a percent as a fractional part of a whole to be of primary importance before working percent problems

(Allinger and Payne 1986). Yet, students who are over reliant on part-whole notions find percentages greater than 100 problematic since in their mind the part cannot exceed the whole (Parker and Leinhardt 1995).

Focusing on this controversy, Moss and Case (1999) report on an experimental curriculum that basically introduced the rational number subconstructs of fraction, decimal, percent in reverse. The curriculum begins rational number instruction with percent in a linear measurement context. It then extends that with instruction with decimals to two places then to three and one places. Finally, instruction leads to fraction notation. Results indicate that students using this curriculum model had a deeper understanding of rational number, less reliance on whole number knowledge, and made more frequent references to proportional reasoning concepts. Moss and Case (1999) found that introducing decimals before fractions led to better educational outcomes than the more traditional sequence of teaching decimals after fractions. In addition, the frequency of the "shorter-is-larger" misconception was higher in countries where fractions were taught before decimals, such as United States and Israel (Nesher and Peled 1986; Resnick et al. 1989) compared to countries like France where decimal fractions expressed in place-value system are taught before other fractions (Resnick et al. 1989; Sackur-Grisvard and Leonard 1985).

9.3 Extending to Probability

Concepts of probability develop in alignment with other ideas in mathematics such as whole number knowledge (counting, addition/subtraction, multiplication/division), fractions (part-whole thinking, equivalent fractions, fraction operations, proportional reasoning), and data (counting/classifying, organization of data, descriptive statistics, data representation and comparison, and sample size). Research into students' ability to compare two probabilities was initiated with work done by Piaget and Inhelder (1975), and other researchers (Fischbein 1975; Fischbein and Gazit 1984; Canizares et al. 1997) have continued the study of students' abilities to reason with probabilities. Comparing probabilities is based on the comparison of two fractions; therefore, proportional reasoning provides the foundation for probabilistic reasoning. Where they differ is that comparing fractions refers to a certain event, whereas comparing (subjective) probabilities involve various amounts of confidence. This degree of uncertainty influences a student's answer because their intuition weighs in on their decision. Intuition is something one considers likely based on instinctive feelings, which may or may not coincide with scientific reasoning, and defines the subjective elements that students assign to probabilities.

In the report *Children's understanding of probability*, Bryant and Nunes (2012) provide a review of four "cognitive demands of understanding probability":

- Understanding randomness
- Working out the sample space
- Comparing and quantifying probabilities
- Understanding correlation (or relationships between events)

The report elaborates on the evidence in each area and highlights studies that are relevant to teaching and learning.

Jones et al. (1997, 1999) evaluated the thinking of 3rd-grade students in an educational setting to create a "Probabilistic Thinking Framework," which describes the levels of reasoning stages related to key constructs (sample space, probability of an event, probability comparisons, and conditional probability). It is interesting to note that these key constructs maintain a striking resemblance to the four cognitive demands made on children when learning probability. Jones and colleagues (1999) also identified four levels of probabilistic thinking. Level 1 is tied to subjective thinking. Level 2 recognizes the transition between subjective thinking and "naïve quantitative thinking" (p. 490). Level 3 engages informal quantitative thinking. Finally, level 4 capitalizes on numerical reasoning. The characteristics within each of the levels of development provide guidance for the design of instructional activities and the selection of instructional strategies that will capitalize on where students are in their reasoning in this sequence of development. Results of this study provide evidence that instruction can influence the learning of probability (Jones et al. 1999).

Way (2003) similarly found three developmental stages of reasoning with probability tasks, along with two distinct transitional stages. In this research, children (of age 4 to 12) were asked to make choices regarding probability tasks and to explain their reasoning. Two types of random number generators were used: Discrete items with up to four colors (numerical form) and spinners with up to four colors (spatial form). Comparisons were drawn within sample space by looking at the colors on one spinner and considering which color is more likely. Comparisons were also made between sample spaces by comparing two spinners and judging with spinner gives the better chance at a specified color (Way 1996, 2003). Key characteristics for each of the age-related stages of development are described in the table below (see Table 1). Children around the age of nine years had built some of the concepts that provide a foundation for probability and demonstrate learning from participation in probability reasoning tasks. At this age, students are likely to benefit from instruction that enables them to make connections among concepts (both within the topic of probability and across mathematical ideas that align with probability) and further expand their early numerical strategies into more sophisticated proportional thinking.

The classification of probabilistic thinking to levels or stages implies a sequential nature that develops slowly over an extended period of time. Batanero and Diaz (2011) argue that the various meanings of probability must be addressed in school mathematics progressively at various levels, beginning with intuitive ideas and a subjective view of probability as a "degree of belief" and building up to formal definitions and approaches. The variability in students' reasoning with probability marks the importance of designing probability tasks and instructional programs in alignment with their cognitive ability and current understandings. It also highlights the significance of formative assessment techniques used to identify and diagnose students' errors and misconceptions as a solid place to start when designing instruction. Hence, teacher training related to probability becomes a concern. Teachers

Table 1 Description of stages of development (from Way 2003)

Stage of development	Age	Key characteristics
1: Non-probabilistic thinking	Average: 5 years 8 months Range: 4 years 3 months to 8 years 2 months	• Minimal understanding of randomness • Reliance on visual comparison • Inability to order likelihood
Transition from non-probabilistic to emergent thinking	Average: 7 years 9 months Range: 5 years 9 months to 11 years 1 months	• Equal number of characteristics from stage 1 and stage 2
2: Emergent probabilistic thinking	Average: 9 years 2 months Range: 6 years 11 months to 12 years 2 months	• Recognition of sample space structure • Ordering of likelihood through visual comparison or estimation of number • Addition/subtraction strategies used in comparison • Concepts of equal likelihood and impossibility
Transition from emergent to quantification	Average: 9 years 5 months Range: 7 years 6 months to 11 years 8 months	• Emergent thinking dominates but there exists evidence of the initiation of quantification
3: Quantification of probability	Average: 11 years 3 months Range: 9 years 1 months to 12 years 7 months	• Numerical comparisons • Doubling and halving • Proportional thinking • Quantification of probability emerging or present

need to be aware of the common errors and misconceptions that students possess and they need to be properly prepared to develop strategies to help students confront them and push their thinking forward (Fischbein and Gazit 1984).

Batanero and Diaz (2011) contend that effective teaching and learning of probability in schools will hinge on proper preparation of teachers. They cite several reasons for the difficulty teachers have in teaching probability. Hill et al. (2005) found that teachers' mathematics preparation and the resulting mathematics knowledge for teaching positively predicted gains in student achievement. Concerns abound that many current teacher-training programs do not sufficiently prepare teachers to teach probability (Franklin and Mewborn 2006). Research indicates that the content knowledge of both pre-service and inservice teachers at elementary and secondary levels is weak. In the late 1980s, the National Center for Research on Teacher Education reported that elementary and secondary teachers were unable to explain their reasoning or why algorithms they used worked (RAND 2003). Instead, their explanations were procedural in nature and lacked conceptual understanding, which is consistent with weak understanding going back to the teachers' own education in schools. A study of 70 secondary mathematics pre-service teachers from two universities found significant content weaknesses (Wilburne and Long 2010), and other research found content knowledge to generally be lacking in conceptual depth (Bryan 2011). Research has documented the struggle of pre-service teachers to identify the source of students' misconceptions and the challenge of finding ways other than the recitation of rules or procedures to eliminate errors and misconceptions (Kilic 2011).

This is consistent with research on inservice teacher knowledge of probability, which was found to be often limited to procedural of formula-based knowledge and to lack conceptual depth. In one study, "most of the teachers primarily had intuitive, informal notions of probability but these later evolved into the classical, frequentist, subjective and mathematical conceptions as they build their conceptual understanding of probability" (Reston 2012, p. 9). Teachers will generally teach as they were taught, so if they are going to incorporate new ideas into their instruction they need to experience those ideas as students. It is already documented that engaging teachers in an inquiry-based approach, which involved strategies that provided for the confrontation of misconceptions, promoted conceptual development of probability and enhanced their pedagogical skills Reston (2012). Albert (2006) argues that subjective probability is often ignored in school curriculum and should play a larger role. The subjective viewpoint is the most general, incorporates the classical and frequentist viewpoints, and expands the definition of probability to include those that cannot be computed and events that cannot be repeated under the same conditions. In his earlier work, Albert (2003) found that college students were marked by overall confusion about the three viewpoints of probability. It is critical to establish different perspectives that enable students to gain confidence in their abilities to reason probabilistically (Albert 2006). According to Shaughnessy (1992), the type of task to be investigated and the type of problem to be solved should drive the viewpoint that is taken in a particular context. Without coherent instruction on probability, students leave the K-12 school system with a jumbled perspective of probability that reflects a mix of the classical, frequentist, and subjectivist viewpoints. Without an understanding of these distinct viewpoints, we will continue to see evidence of children and adults inappropriately applying these models of probability.

What exactly should an ecologically rational pedagogy for mathematics education look like, including the development of probabilistic thinking? In general, educational practices should work to bootstrap from these privileged representations (rather than fight them) and built towards a more inclusive and comprehensive model of mathematics knowledge. In summary, we suggest the following educational plan would be more effective and less difficult (for both students and teachers). *Start with frequencies*. This is the natural starting point and should be recognized as such, returning to this starting point as necessary when building up to some more advanced concept. *Introduce percentages and decimals before fractions*. Given the partitioning nature of these numerical formats, relative to whole numbers, it makes more sense to start with formats that use standard units of partitioning first (i.e., percentages can be thought of as 100 parts; decimals can be thought of as successive partitions of 10). This can actually help to lay the groundwork of the much more partition-complex situation represented by fractions. *Explain percentages, decimals, and fractions in terms of how they differ from frequencies*. Teachers should expect that many students will adopt frequentist interpretations of these numerical formats. Rather than pretend that those are arbitrarily or definitively wrong, without any explanation, teachers should be allowed and encouraged to address common in an explicit and transparent way. This can lead to student not just seeing the surface similarities between frequencies and other formats, but also understanding that there are further properties of non-frequency formats. *Finally, probabilities should be introduced in a similar manner*. Probabilities based on a frequentist interpretation are valid (just like, in some contexts frequency-based percentages, decimals and fractions can exist). This can be recognized, acknowledged as a real perspective, but then students can be progressively moved to a more extensive and complete understanding of what probabilities can be (either frequentist or subjectivist interpretations). Accomplishing this can lead to students who not only engage in correct probabilistic reasoning, but who understand why and how they did so.

References

Albert, J. (2003). Conceptions of probability of college students. *American Statistician*, *57*, 37–45.

Albert, J. (2006). Interpreting probabilities and teaching the subjectivist viewpoint. In G. Burrill (Ed.), *Thinking and reasoning about data and chance: 68th NCTM yearbook, 2006* (pp. 417–433). Reston: National Council of Teachers of Mathematics.

Allinger, G. D., & Payne, J. N. (1986). Estimation and mental arithmetic with percent. In H. L. Schoen (Ed.), *Estimation and mental computation* (pp. 141–155). Reston: National Council of Teachers of Mathematics.

Ansari, D. (2008). Effects of development and enculturation on number representation in the brain. *Nature Reviews. Neuroscience*, *9*(4), 278–291. doi:10.1038/nrn2334.

Ansari, D., Garcia, N., Lucas, E., Hamon, K., & Dhital, B. (2005). Neural correlates of symbolic number processing in children and adults. *Neuroreport*, *16*, 1769–1773.

Ball, D. (1993). Halves, pieces and twoths: constructing and using representational contexts in teaching fractions. In T. P. Carpenter, E. Fennema, & T. A. Romberg (Eds.), *Rational numbers: an integration of research* (pp. 157–195). Hillsdale: Erlbaum.

Barbey, A. K., & Sloman, S. A. (2007). Base-rate respect: from ecological rationality to dual processes. *Behavioral and Brain Sciences*, *30*(3), 241–254.

Batanero, C., & Diaz, C. (2011). Training school teachers to teach probability: reflections and challenges. *Chilean Journal of Statistics*. Retrieved August 2012 from http://chjs.deuv.cl/iFirst_art/ChJS010202.pdf.

Behr, M. J., & Post, T. R. (1988). Teaching rational number and decimal concepts. In T. R. Post (Ed.), *Teaching mathematics in grades K-8* (pp. 201–248). Boston: Allyn and Bacon.

Behr, M. J., Wachsmuth, I., Post, T. R., & Lesh, R. (1984). Order and equivalence of rational numbers: a clinical teaching experiment. *Journal for Research in Mathematics Education*, *15*, 323–341.

Beran, M. J. (2007). Rhesus monkeys (Macaca mulatta) enumerate sequentially presented sets of items using analog numerical representations. *Journal of Experimental Psychology. Animal Behavior Processes*, *33*, 42–54.

Beran, M. J. (2008). The evolutionary and developmental foundations of mathematics. *PLoS Biology*, *6*(2), e19. doi:10.1371/journal.pbio.0060019.

Beran, M. J., & Rumbaugh, D. M. (2001). "Constructive" enumeration by chimpanzees (Pan troglodytes). *Animal Cognition*, *4*, 81–89.

Boysen, S. T., & Berntson, G. G. (1989). Numerical competence in a chimpanzee (Pan troglodytes). *Journal of Comparative Psychology*, *103*, 23–31.

Brannon, E. M., & Terrace, H. S. (2000). Representation of the numerosities 1–9 by rhesus macaques (Macaca mulatta). *Journal of Experimental Psychology. Animal Behavior Processes*, *26*, 31–49.

Brannon, E. M., & Terrace, H. S. (2002). The evolution and ontogeny of ordinal numerical ability. In M. Bekoff, C. Allen, & G. Burghardt (Eds.), *The cognitive animal: empirical and theoretical perspectives on animal cognition* (pp. 197–204). Cambridge: MIT Press.

Brase, G. L. (2002a). Which statistical formats facilitate what decisions? The perception and influence of different statistical information formats. *Journal of Behavioral Decision Making*, *15*(5), 381–401.

Brase, G. L. (2002b). Ecological and evolutionary validity: comments on Johnson-Laird, Legrenzi, Girotto, Legrenzi, and Caverni's (1999) mental-model theory of extensional reasoning. *Psychological Review*, *109*(4), 722–728.

Brase, G. L. (2002c). "Bugs" built into the system: how privileged representations influence mathematical reasoning across the lifespan. *Learning and Individual Differences*, *12*, 391–409.

Brase, G. L. (2007). Omissions, conflations, and false dichotomies: conceptual and empirical problems with the Barbey & Sloman account. *Behavioral and Brain Sciences*, *30*(3), 258–259.

Brase, G. L. (2008). Frequency interpretation of ambiguous statistical information facilitates Bayesian reasoning. *Psychonomic Bulletin & Review*, *15*(2), 284–289.

Brase, G. L. (2009). Pictorial representations in statistical reasoning. *Applied Cognitive Psychology*, *23*(3), 369–381.

Brown, D. (1991). *Human universals*. New York: McGraw-Hill.

Brown, J. S., & Burton, R. R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science*, *2*(2), 155–192.

Bryan, T. J. (2011). The conceptual knowledge of preservice secondary mathematics teachers: how well do they know the subject matter they will teach? *Issues in the Undergraduate Preparation of Teachers: The Journal*, *1*.

Bryant, P., & Nunes, T. (2012). *Children's understanding of probability: a literature review*. London: Nuffield Foundation.

Butterworth, B. (1999). *The mathematical brain*. London: Macmillan.

Canizares, M. J., Batareno, C., Serrano, L., & Ortiz, J. J. (1997). Subjective elements in children's comparison of probabilities. In E. Pehkonen (Ed.), *Proceedings of the 21st conference of the international group for the psychology of mathematics education (PME)*. University of Helsinki.

Cantlon, J. F., & Brannon, E. M. (2006). Shared system for ordering small and large numbers in monkeys and humans. *Psychological Science*, *17*, 401–406.

Cantlon, J. F., & Brannon, E. M. (2007). Basic math in monkeys and college students. *PLoS Biology*, *5*(12), e328. doi:10.1371/journal.pbio.0050328.

Cantlon, J. F., Brannon, E. M., Carter, E. J., & Pelphrey, K. A. (2006). Functional imaging of numerical processing in adults and 4-y-old children. *PLoS Biology*, *4*, e125.

Carraher, D. W. (1996). Learning about fractions. In L. P. Steffe, P. Nesher, G. Goldin, P. Cobb, & B. Greer (Eds.), *Theories of mathematical learning* (pp. 241–266). Hillsdale: Erlbaum.

Casasanto, D. (2005). Crying "Whorf". *Science*, *307*, 1721–1722.

Casscells, W., Schoenberger, A., Grayboys, T., & Graboys, T. B. (1978). Interpretation by physicians of clinical laboratory results. *The New England Journal of Medicine*, *299*(18), 999–1000. doi:10.1056/NEJM197811022991808.

Castillo-Garsow, C. W. (2010). *Teaching the Verhulst model: a teaching experiment in covariational reasoning and exponential growth*. PhD thesis, Arizona State University, Tempe, AZ.

Clements, M. A., & Del Campo, G. (1990). How natural is fraction knowledge? In L. P. Steffe & T. Wood (Eds.), *Transforming children's mathematics education: international perspectives* (pp. 181–188). Hillsdale: Erlbaum.

Cohen Kadosh, R., Cohen Kadosh, K., Kaas, A., Henik, A., & Goebel, R. (2007). Notation-dependent and -independent representations of numbers in the parietal lobes. *Neuron*, *53*(2), 307–314. doi:10.1016/j.neuron.2006.12.025.

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

D'Ambrosio, B. S., & Mewborn, D. S. (1994). Children's constructions of fractions and their implications for classroom instruction. *Journal of Research in Childhood Education*, *8*, 150–161.

De Cruz, H. (2006). Why are some numerical concepts more successful than others? An evolutionary perspective on the history of number concepts. *Evolution and Human Behavior*, *27*(4), 306–323.

Dehaene, S. (1997). *The number sense: how the mind creates mathematics*. New York: Oxford University Press.

Dehaene, S., Izard, V., Pica, P., & Spelke, E. (2006). Core knowledge of geometry in an Amazonian indigene group. *Science*, *311*(5759), 381–384. doi:10.1126/science.1121739.

Diester, I., & Nieder, A. (2007). Semantic associations between signs and numerical categories in the prefrontal cortex. *PLoS Biology*, *5*, e294.

Erlwanger, S. H. (1973). Benny's conception of rules and answers in IPI mathematics. *The Journal of Children's Mathematical Behavior*, *1*(2), 7–26.

Everett, D. L. (2005). Cultural constraints on grammar and cognition in Pirahã. Another look at the design features of human language. *Current Anthropology*, *46*(4), 621–646.

Feigenson, L., Carey, S., & Hauser, M. D. (2002). The representations underlying infants' choice of more: object files versus analog magnitudes. *Psychological Science*, *13*, 150–156.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314. doi:10.1016/j.tics.2004.05.002.

Fischbein, E. (1975). *The intuitive sources of probabilistic thinking in children*. Dordrecht: Reidel.

Fischbein, E., & Gazit, A. (1984). Does the teaching of probability improve probabilistic intuitions? *Educational Studies in Mathematics*, *15*, 1–24.

Frank, M. C., Everett, D. L., Fedorenko, E., & Gibson, E. (2008). Number as a cognitive technology: evidence from Pirahã language and cognition. *Cognition*, *108*(3), 819–824. doi:10.1016/j.cognition.2008.04.007.

Franklin, C., & Mewborn, D. (2006). The statistical education of PreK-12 teachers: a shared responsibility. In *NCTM 2006 yearbook: thinking and reasoning with data and chance* (pp. 335–344).

Fuson, K. C. (1990). Conceptual structures for multiunit numbers: implications for learning and teaching multidigit addition, subtraction, and place value. *Cognition and Instruction*, *7*(4), 343–403.

Gallistel, C. R. (1990). *The organization of learning*. Cambridge: MIT Press.

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, *44*, 43–74.

Gallistel, C. R., Gelman, R., & Cordes, S. (2006). The cultural and evolutionary history of the real numbers. In S. C. Levinson & J. Pierre (Eds.), *Evolution and culture: a Fyssen Foundation symposium* (pp. 247–274). Cambridge: MIT Press.

Garner, W. R., Hake, H. W., & Eriksen, C. W. (1956). Operationism and the concept of perception. *Psychological Review*, *63*(3), 149–159.

Gay, A. S., & Aichele, D. B. (1997). Middle school students' understanding of number sense related to percent. *School Science and Mathematics*, *97*, 27–36.

Geary, D. C. (1995). Reflections of evolution and culture in children's cognition: implications for mathematical development and instruction. *The American Psychologist*, *50*(1), 24–37.

Geary, D. C. (2007). An evolutionary perspective on learning disability in mathematics. *Developmental Neuropsychology*, *32*(1), 471–519.

Geary, D. C., Hoard, M. K., & Hamson, C. O. (1999). Numerical and arithmetical cognition: patterns of functions and deficits in children at risk for a mathematical disability. *Journal of Experimental Child Psychology*, *74*(3), 213–239.

Gelman, R., & Meck, E. (1992). Early principles aid initial but not later conceptions of number. In J. Bideaud, C. Meljac, & J. P. Fischer (Eds.), *Pathways to number* (pp. 171–189, and addendum to book). Hillsdale: Erlbaum.

Gigerenzer, G. (1993). The superego, the ego, and the id in statistical reasoning. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences* (pp. 311–339). Hillsdale: Erlbaum

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: frequency formats. *Psychological Review*, *102*(4), 684–704. doi:10.1037/0033-295X.102.4.684.

Gigerenzer, G., Swijtink, Z. G., Porter, T. M., Daston, L., Beatty, J., & Kruger, L. (1989). *The empire of chance: how probability changed science and everyday life*. New York: Cambridge University Press.

Gigerenzer, G., Todd, P. M., & Group T. A. B. C. R. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Gilmore, C. K., McCarthy, S. E., & Spelke, E. S. (2007). Symbolic arithmetic knowledge without instruction. *Nature*, *447*(7144), 589–591. doi:10.1038/nature05850.

Girotto, V., & Gonzalez, M. (2001). Solving probabilistic and statistical problems: a matter of information structure and question form. *Cognition*, *78*(3), 247–276.

Glasgow, R., Ragan, G., & Fields, W. M. (2000). The decimal dilemma. *Teaching Children Mathematics*, *7*(2), 89.

Glassman, M. (1996). The argument for constructivism. *The American Psychologist*, *51*(3), 264–265.

Gordon, P. F. (2004). Numerical cognition without words: evidence from Amazonia. *Science*, *306*, 496–499.

Gordon, P. (2005). Author's response to "Crying Whorf". *Science*, *307*, 1722.

Gould, P. (2005). Year 6 students' methods of comparing the size of fractions. In P. Clarkson, A. Downton, D. Gronn, M. Horne, A. McDonough, R. Pierce, & A. Roche (Eds.), *Building connections: research, theory and practice*. *Proceedings of the annual conference of the Mathematics Education Research Group of Australasia* (pp. 393–400). Sydney: MERGA.

Hacking, I. (1975). *The emergence of probability: a philosophical study of early ideas about probability induction and statistical inference*. Cambridge: Cambridge University Press.

Hall, K. (1975). Gaps in grammar and culture. In M. D. Kinkade, K. L. Hale, & O. Werner (Eds.), *Linguistics and anthropology: in honor of C.F. Voegelin* (pp. 295–315). Lisse: Peter De Riddder Press.

Harel, G., & Sowder, L. (2005). Advanced mathematical-thinking at any age: its nature and its development. *Mathematical Thinking and Learning*, *7*, 27–50.

Hart, K. M. (1981). *Children's understanding of mathematics: 11–16. The CSMS Mathematics Team*. London: Murray.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology. General*, *108*(3), 356–388.

Hasher, L., & Zacks, R. T. (2002). Frequency processing: a twenty-five year perspective. In P. Sedlmeier & T. Betsch (Eds.), *Frequency processing and cognition* (pp. 21–36). Oxford: Oxford University Press.

Hiebert, J. (1984). Children's mathematics learning: the struggle to link form and understanding. *The Elementary School Journal*, *84*(5), 497–511.

Hiebert, J., & Wearne, D. (1985). A model of students' decimal computation procedures. *Cognition and Instruction*, *2*(3&4), 175–205.

Hiebert, J., & Wearne, D. (1986). Procedures over concepts: the acquisition of decimal number knowledge. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: the case of mathematics*.

Hiebert, J., Wearne, D., & Taber, S. (1991). Fourth graders' gradual construction of decimal fractions during instruction using different physical representations. *The Elementary School Journal*, *91*, 321–341.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *42*(2), 371–406.

Hintzman, D. L., & Stern, L. D. (1978). Contextual variability and memory for frequency. *Journal of Experimental Psychology. Human Learning and Memory*, *4*(5), 539–549.

Hoffrage, U., Gigerenzer, G., Krauss, S., & Martignon, L. (2002). Representation facilitates reasoning: what natural frequencies are and what they are not. *Cognition*, *84*(3), 343–352. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12044739.

Hoffrage, U., Kurzenhauser, S., & Gigerenzer, G. (2005). Understanding the results of medical tests: why the representation of statistical information matters. In R. Bibace, J. D. Laird, K. L. Noller, & J. Valsiner (Eds.), *Science and medicine in dialogue: thinking through particulars and universals* (pp. 83–98). Westport: Praeger Publishers/Greenwood Publishing Group.

Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Communicating statistical information. *Science*, *290*(5500), 2261–2262.

Inoue, S., & Matsuzawa, T. (2007). Working memory of numerals in chimpanzees. *Current Biology*, *17*(23), R1004–R1005. doi:10.1016/j.cub.2007.10.027.

Izard, V., Dehaene-Lambertz, G., & Dehaene, S. (2008). Distinct cerebral pathways fro object identity and number in human infants. *PLoS Biology*, *6*, e11.

Jacob, S. N., & Nieder, A. (2009). Notation-independent representation of fractions in the human parietal cortex. *The Journal of Neuroscience*, *29*(14), 4652–4657. doi:10.1523/JNEUROSCI.0651-09.2009.

Jones, G., Langrall, C., Thornton, C., & Mogill, T. (1997). A framework for assessing and nurturing young children's thinking in probability. *Educational Studies in Mathematics*, *32*, 101–125.

Jones, G., Langrall, C., Thornton, C., & Mogill, T. (1999). Using probabilistic thinking to inform instruction. *Journal for Research in Mathematics Education*, *30*(5), 487–519.

Jones, G. A., & Thornton, C. A. (2005). An overview of research into the learning and teaching of probability. In G. A. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 65–92). New York: Springer.

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus, and Giroux.

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

Kent, L. B., Arnosky, J., & McMonagle, J. (2002). Using representational contexts to support multiplicative reasoning. In B. Litwiller & G. Bright (Eds.), *Making sense of fractions, ratios, and proportions: 2002 yearbook* (pp. 145–152). Reston: National Council of Teachers of Mathematics.

Kerslake, D. (1986). *Fractions: children's strategies and errors*. London: NFER-Nelson.

Kieren, T. E. (1976). On the mathematical, cognitive, and instructional foundations of rational numbers. In R. Lesh (Ed.), *Number and measurement: papers from a research workshop* (pp. 101–144). Columbus: ERIC Center for Science, Mathematics, and Environmental Education.

Kieren, T. E. (1980). The rational number construct: its elements and mechanisms. In T. Kieran (Ed.), *Recent research on number learning* (pp. 125–149). Columbus: ERIC-SMEAC.

Kieren, T. E. (1988). *Personal knowledge of rational numbers: its intuitive and formal the middle grades* (pp. 162–181). Reston: Author.

Kilic, H. (2011). Preservice secondary mathematics teachers' knowledge of students. *Turkish Online Journal of Qualitative Inquiry*, *2*(2).

Kleiter, G. (1994). Natural sampling: rationality without base rates. In G. H. Fischer & D. Laming (Eds.), *Contributions to mathematical psychology, psychometrics, and methodology* (pp. 375–388). New York: Springer.

Lachance, A., & Confrey, J. (2001). Helping students build a path of understanding from ratio and proportion to decimal notation. *The Journal of Mathematical Behavior*, *20*(4), 503–526.

Lamon, S. J. (2001). Presenting and representing: from fractions to rational numbers. In A. Cuoco & F. R. Curcio (Eds.), *The role of representation in school mathematics* (pp. 146–165). Reston: National Council of Teachers of Mathematics.
Lesh, R., Landau, M., & Hamilton, E. (1983). Conceptual models in applied mathematical problem solving research. In R. Lesh & M. Landau (Eds.), *Acquisition of mathematics concepts & processes* (pp. 263–343). New York: Academic Press.
Lipton, J. S., & Spelke, E. S. (2003). Origins of number sense. Large-number discrimination in human infants. *Psychological Science*, *14*(5), 396–401. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12930467.
Mack, N. K. (1990). Learning fractions with understanding: building on informal knowledge. *Journal for Research in Mathematics Education*, *21*(1), 16–33.
Mack, N. K. (1993). Learning rational numbers with understanding: the case of informal knowledge. In T. P. Carpenter, E. Fennema, & T. A. Romberg (Eds.), *Rational numbers: an integration of research* (pp. 85–106). Hillsdale: Erlbaum.
Mack, N. K. (1995). Confounding whole-number and fraction concepts when building on informal knowledge. *Journal for Research in Mathematics Education*, *26*, 422–441.
Markovits, Z., & Hershkowitz, R. (1997). Relative and absolute thinking in visual estimation processes. *Educational Studies in Mathematics*, *32*, 29–47.
Matsuzawa, T. (1985). Use of numbers by a chimpanzee. *Nature*, *315*, 57–59.
McCrink, K., & Wynn, K. (2004). Large-number addition and subtraction by 9-month-old infants. *Psychological Science*, *15*(11), 776–781. doi:10.1111/j.0956-7976.2004.00755.x.
McNeil, N. M., & Alibali, M. W. (2005). Why won't you change your mind? Knowledge of operational patterns hinders learning and performance on equations. *Child Development*, *76*, 883–899.
Mix, K. S., Huttenlocher, J., & Levine, S. C. (2002). *Quantitative development in infancy and early childhood*. New York: Oxford University Press.
Mix, K. S., & Sandhofer, C. M. (2007). Do we need a number sense? In M. J. Roberts (Ed.), *Integrating the mind*. Hove: Psychology Press.
Moloney, K., & Stacey, K. (1997). Changes with age in students' conceptions of decimal notation. *Mathematics Education Research Journal*, *9*(1), 25–38.
Moss, J., & Case, R. (1999). Developing children's understanding of the rational numbers: a new model and an experimental curriculum. *Journal for Research in Mathematics Education*, *30*, 122–147.
Nesher, P., & Peled, I. (1986). Shifts in reasoning. *Educational Studies in Mathematics*, *17*, 67–79.
Ni, Y. (2005). Teaching and learning fraction and rational numbers: the origins and implications of whole number bias. *Educational Psychologist*, *40*(1), 27–52. doi:10.1207/s15326985ep4001_3. (776109343).
Over, D. E., & Green, D. W. (2001). Contingency, causation, and adaptive inference. *Psychological Review*, *108*(3), 682–684.
Parker, M., & Leinhardt, G. (1995). Percent: a privileged proportion. *Review of Educational Research*, *65*(4), 421–481.
Payne, J. N., & Allinger, G. D. (1984). *Insights into teaching percent to general mathematics students*. Unpublished manuscript, Montana State University, Department of Mathematical Sciences, Bozeman.
Piaget, J., & Inhelder, B. (1975). *The origin of the idea of chance in children*. London: Routledge & Kegan Paul. (Original work published 1951).
Piazza, M., Pinel, P., Le Bihan, D., & Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron*, *53*(2), 293–305. doi:10.1016/j.neuron.2006.11.022.
RAND Mathematics Study Panel [RAND MSP] (2003). *Mathematical proficiency for all students: towards a strategic development program in mathematics education*. Santa Monica: RAND Corporation. (MR-1643.0-OERI).
Real, L. (1991). Animal choice behavior and the evolution of cognitive architecture. *Science*, *253*, 980–986.

Real, L., & Caraco, T. (1986). Risk and foraging in stochastic environments: theory and evidence. *Annual Review of Ecology and Systematics*, *17*, 371–390.

Resnick, L. B., Nesher, P., Leonard, F., Magone, M., Omanson, S., & Peled, I. (1989). Conceptual bases of arithmetic errors: the case of decimal fractions. *Journal for Research in Mathematics Education*, *20*, 8–27.

Reston, E. (2012). Exploring inservice elementary mathematics teachers' conceptions of probability through inductive teaching and learning methods. In *12th international congress on mathematics education*. Seoul, Korea.

Risacher, B. F. (1992). Knowledge growth of percent during the middle school years. *Diss. Abstr. Int.*, *54*(03), 853A.

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: an iterative process. *Journal of Educational Psychology*, *93*(2), 346–362.

Rivera, S. M., Reiss, A. L., Eckert, M. A., & Menon, V. (2005). Developmental changes in mental arithmetic: evidence for increased functional specialization in the left inferior parietal cortex. *Cerebral Cortex*, *15*, 1779–1790.

Sackur-Grisvard, C., & Leonard, F. (1985). Intermediate cognitive organizations in the process of learning a mathematical concept: the order of positive decimal numbers. *Cognition and Instruction*, *2*(2), 157–174.

Schmitt, D. P., & Pilcher, J. J. (2004). Evaluating evidence of psychological adaptation: how do we know one when we see one? *Psychological Science*, *15*(10), 643–649. doi:10.1111/j.0956-7976.2004.00734.x.

Sedlmeier, P., Hertwig, R., & Gigerenzer, G. (1998). Are judgments of the positional frequencies of letters systematically biased due to availability? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *24*(3), 754–770. doi:10.1037/0278-7393.24.3.754.

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.

Shepard, R. N. (1984). Ecological constraints on internal representation: resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological Review*, *91*, 417–447.

Siegler, R. (2000). The rebirth of children's learning. *Child Development*, *71*(1), 26–35.

Silver, E. A. (1986). Using conceptual and procedural knowledge: a focus on relationships. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: the case of mathematics* (pp. 181–198). Hillsdale: Erlbaum.

Simon, H. A. (1956). Rational choice and the structure of environments. *Psychological Review*, *63*, 129–138.

Sloman, S. a., Over, D., Slovak, L., & Stibel, J. M. (2003). Frequency illusions and other fallacies. *Organizational Behavior and Human Decision Processes*, *91*(2), 296–309. doi:10.1016/S0749-5978(03)00021-9.

Sophian, C. (2000). From objects to quantities: developments in preschool children's judgments about aggregate amount. *Developmental Psychology*, *36*(6), 724–730.

Sowder, J., Armstrong, B., Lamon, S., Simon, M., Sowder, L., & Thompson, A. (1998). Educating teachers to teach multiplicative structures in the middle grades. *Journal of Mathematics Teacher Education*, *1*, 127–155.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*(1), 89–96. doi:10.1111/j.1467-7687.2007.00569.x.

Staddon, J. F. R. (1988). Learning as inference. In R. C. Bolles & M. D. Beecher (Eds.), *Evolution and learning* (pp. 59–77). Hillsdale: Erlbaum.

Steinle, V. (2004). *Changes with age in students' misconceptions of decimals numbers*. University of Melbourne.

Steinle, V., & Stacey, K. (2004a). A longitudinal study of students understanding of decimal notation: an overview and refined results. In I. Putt, R. Faragher, & M. McLean (Eds.), *Proceedings of the 27th annual conference of the Mathematics Education Research Group of Australasia* (Vol. 2, pp. 541–548). Townsville: MERGA.

Steinle, V., & Stacey, K. (2004b). Persistence of decimal misconceptions and readiness to move to expertise. In *Proceedings of the 28th conference of the international group for the psychology of mathematics education* (Vol. 4, pp. 225–232). Bergen: PME.

Stephens, D., & Krebs, J. (1986). *Foraging theory*. Princeton: Princeton University Press.

Sternberg, R. J., & Grigorenko, E. L. (2001a). Unified psychology. *The American Psychologist*, *56*(12), 1069–1079.

Sternberg, R. J., & Grigorenko, E. I. (2001b). The (mis)organization of psychology. *APS Observer*, *14*(1).

Streefland, L. (1991). *Fractions in realistic mathematics education*. Dordrecht: Kluwer Academic.

Thompson, P. W. (1994). The development of the concept of speed and its relationship to concepts of rate. In G. Harel & J. Confrey (Eds.), *The development of multiplicative reasoning in the learning of mathematics* (pp. 181–234). Albany: SUNY Press.

Thompson, P. W. (2004). *Fractions and multiplicative reasoning* (pp. 95–114). Reston: National Council of Teachers of Mathematics.

Thompson, P. W., & Saldanha, L. A. (2004). Fractions and multiplicative reasoning. In J. Kilpatrick, G. Martin, & D. Schifter (Eds.), *Research companion to the principles and standards for school mathematics* (pp. 95–114). Reston: National Council of Teachers of Mathematics.

Thompson, P. W., & Thompson, A. G. (1992). *Images of rate*. San Francisco, CA.

Thompson, P. W., & Thompson, A. G. (1994). Talking about rates conceptually, part I: a teacher's struggle. *Journal for Research in Mathematics Education*, *25*, 279–303.

Thompson, A. G., & Thompson, P. W. (1996). Talking about rates conceptually, part II: mathematical knowledge for teaching. *Journal for Research in Mathematics Education*, *27*, 2–24.

Thornton, M. C., & Fuller, R. G. (1981). How do college students solve proportion problems? *Journal of Research in Science Teaching*, *18*(4), 335–340.

Tomonaga, M., & Matsuzawa, T. (2002). Enumeration of briefly presented items by the chimpanzee (Pan troglodytes) and humans (Homo sapiens). *Animal Learning & Behavior*, *30*, 143–157.

Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: evolutionary psychology and the generation of culture* (pp. 19–136). Oxford: Oxford University Press.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, *76*, 105–110.

Tversky, A., & Kahneman, D. (1973). Availability: a heuristic for judging frequency and probability. *Cognitive Psychology*, *4*, 207–232.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*, 453–458.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. *Psychological Review*, *90*, 293–315.

Vamvakoussi, X., & Vosniadou, S. (2004). Understanding the structure of rational numbers: a conceptual change approach. *Learning and Instruction*, *14*, 453–467. Special issue. L. Verschaffel & S. Vosniadou (Eds.), Conceptual change in mathematics learning and teaching.

Van Marle, K., & Wynn, K. (2011). Tracking and quantifying objects and non-cohesive substances. *Developmental Science*, *14*(3), 502–515. doi:10.1111/j.1467-7687.2010.00998.x.

Vergnaud, G. (1994). The multiplicative conceptual field: what and why? In G. Harel & J. Confrey (Eds.), *The development of multiplicative reasoning in the learning of mathematics* (pp. 41–59). Albany: SUNY Press.

Wang, X. T. (1996a). Domain-specific rationality in human choices: violations of utility axioms and social contexts. *Cognition*, *60*(1), 31–63. doi:10.1016/0010-0277(95)00700-8.

Wang, X. T. (1996b). Framing effects: dynamics and task domains. *Organizational Behavior and Human Decision Processes*, *68*, 145–157.

Wang, X. T., & Johnston, V. S. (1995). Perceived social context and risk preference: a re-examination of framing effects in a life-death decision problem. *Journal of Behavioral Decision Making*, *8*, 279–293.

Wang, X. T., Simons, F., & Brédart, S. (2001). Social cues and verbal framing in risky choice. *Journal of Behavioral Decision Making*, *14*(1), 1–15. doi:10.1002/1099-0771(200101)14:1<1::AID-BDM361>3.3.CO;2-E.

Way, J. (1996). Children's strategies for comparing two types of random generators. In L. Puig & A. Guitierrez (Eds.), *Proceedings of the twentieth international conference for the psychology of mathematics education*, Valencia, Spain (Vol. 4, pp. 419–426).

Way, J. (2003). The development of young children's notions of probability. *CERME*, *3*, Italy.

Wilburne, J. M., & Long, M. (2010). Secondary pre-service teachers' content knowledge for state assessments: implications for mathematics education programs. *IUMPST: The Journal*, *1*. (Content knowledge). www.k-12prep.math.ttu.edu.

Woodward, J., Howard, L., & Battle, R. (1997). *Learning subtraction in the third and fourth grade: what develops over time*. (Tech. Report 97-2). University of Puget Sound, Tacoma.

Wynn, K. (1998a). Numerical competence in infants. In C. Donlan (Ed.), *Studies in developmental psychology*. *The development of mathematical skills*.

Wynn, K. (1998b). An evolved capacity for number. In D. Cummins Dellarosa & C. Allen (Eds.), *The evolution of mind* (pp. 107–126). New York: Oxford University Press.

Xu, F. (2003). Numerosity discrimination in infants: evidence for two systems of representations. *Cognition*, *89*(1), B15–B25. doi:10.1016/S0010-0277(03)00050-7.

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(13), 5012–5015. doi:10.1073/pnas.0704450105.

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*(1), B1–B11.

Zazkis, R., & Chernoff, E. (2008). What makes a counterexample exemplary? *Educational Studies in Mathematics*, *68*(3), 195–208.

Zhu, L., & Gigerenzer, G. (2006). Children can solve Bayesian problems: the role of representation in mental computation. *Cognition*, *98*(3), 287. doi:10.1016/j.cognition.2004.12.003.

The Interplay Among Knowledge, Cognitive Abilities and Thinking Styles in Probabilistic Reasoning: A Test of a Model

Francesca Chiesi and Caterina Primi

Abstract Stanovich et al. (Adv. Child Dev. Behav. 36, 251–285, 2008) outlined how people can reach a correct solution when a task besides the normative solution elicits competing response options that are intuitively compelling. First of all, people have to possess the relevant rules, procedures, and strategies derived from past learning experiences, called *mindware* (Perkins, Outsmarting IQ: the emerging science of learnable *intelligence*, Free Press, New York, 1995). Then they have to recognise the need to use and to inhibit competing responses. Starting from this assumption, Stanovich and colleagues developed a taxonomy of thinking errors that builds on the dual-process theories of cognition.

The present chapter presents a set of experiments designed to test the Stanovich and colleagues' model inside probabilistic reasoning. Since rules concerned with probabilistic reasoning (i.e. the *mindware* in Stanovich and colleagues' terms) are learned and consolidated through education, we carried on the researches with students of different grade levels. In particular, we assessed the role of the *mindware gap* (i.e. missing knowledge), taking into account individual differences in cognitive ability and thinking dispositions, and superstitious thinking as *contaminated* mindware (Study 1). Then, we conducted a set of experiments (Study 2) in order to investigate the *override failure* (i.e. the failure in inhibiting intuitive competing responses) in which participants were instructed to reason on the basis of logic or provided with example of logical vs. intuitive solutions of the same task. In this way, we aimed at stressing the need to apply the rules.

Our results provide support for the claim that the *mindware* plays an important role in probabilistic reasoning independent of age. Moreover, we found that cognitive capacity increases reasoning performance only if individuals possess the necessary knowledge about normative rules. Finally, superstitious beliefs seem to have a detrimental effect on reasoning. The overall findings offer some cues to cross the bridge from a psychological approach to an educational approach.

F. Chiesi (✉) · C. Primi
Department of Neuroscience, Psychology, Drug Research, and Child's Health (NEUROFARBA), Section of Psychology, University of Florence, Via di San Salvi 12, Padiglione 26, 50135 Florence, Italy
e-mail: francesca.chiesi@unifi.it

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_11,

According to dual-process theories, mental functioning can be characterized by two different types of process which have different functions and different strengths and weaknesses (e.g. Brainerd and Reyna 2001; Epstein 1994; Evans and Over 1996; Stanovich 1999).

To demonstrate the role of the two types of process in reasoning, consider the following example (Chiesi et al. 2011). Imagine that in order to win a prize you have to pick a red marble from one of two urns (Urn A and B). Urn A contains 20 red and 80 blue marbles, and Urn B contains 1 red and 9 blue marbles. When you respond to the task, you can simply rely on the feeling/intuition that it is preferable to pick from the urn with more red marbles (i.e. the favourable events). In this case, you are using Type 1 processes (which are sometimes called *heuristic*—see, e.g. Evans 2006; Klaczynski 2004) that they are considered to be autonomous because their execution is rapid and mandatory when the triggering stimuli are encountered, they do not require much cognitive effort, and they can operate in parallel. Type 1 processing is the default because it is cognitively economical, and people "feel" intuitively that heuristic responses are correct (Epstein 1994; Thompson 2009). Indeed, as in the example, Type 1 processing often leads to normative correct responses (e.g. Evans 2003; Stanovich and West 1999).

You could also respond to the task comparing the ratio of winning marbles in each urn (20 % vs. 10 %) which requires some time, mental effort and computations. In this case, you are using Type 2 processes, which are relatively slow and computationally expensive, available for conscious awareness, serial, and often language based. In this example, both processes cue the normatively correct answer (that is, Urn A).

On the other hand, it is possible to set up a task where Type 1 and Type 2 reasoning cue different responses. For example, if you can choose between picking a marble from an urn containing 10 red and 90 blue marbles, or from an urn containing 2 red and 8 blue marbles, the feeling/intuition that it is preferable to pick from the urn with more favourable events results in a normatively incorrect choice.

When Type 1 and Type 2 processes do not produce the same output, Type 1 process usually cues responses that are normatively incorrect and, according to dual-process theorists (e.g. Stanovich 1999) one of the most critical functions of Type 2 process in these cases is to interrupt and override Type 1 processing. However, this does not always happen. In the case of a conflict between intuitions and normative rules, even educated adults will predominantly produce heuristic responses (e.g. Klaczynski 2001).

To account for this finding, Stanovich and West (2008; see also Stanovich et al. 2008) suggested that the Type 2 override process crucially depends on whether people detect the conflict between their intuitions and their knowledge about relevant normative rules. These rules, procedures, and strategies derived from past learning experiences have been referred to as *mindware* (Perkins 1995). The concept of mindware was adopted by Stanovich and colleagues (Stanovich and West 2008) in their recent model of the role of knowledge in producing normative responses to reasoning problems (Fig. 1). If the relevant mindware can be retrieved and used, alternative responses become available to engage in the override of the intuitive

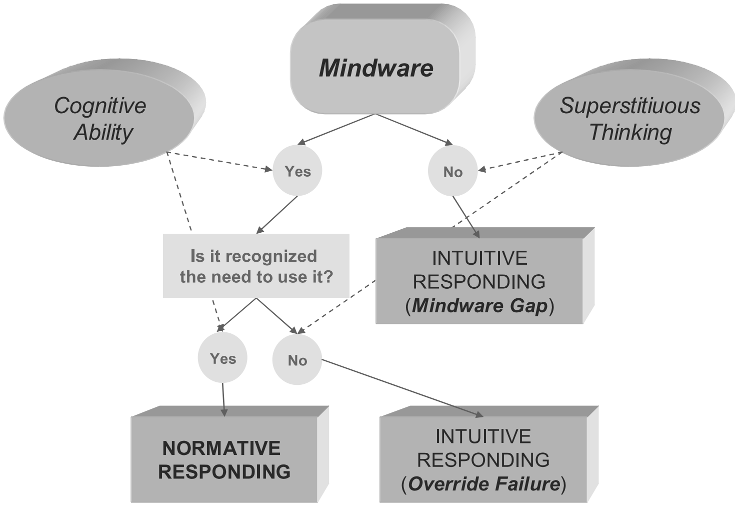


Fig. 1 Simplified representation of Stanovich et al.'s (2008) model on normative reasoning and thinking errors

compelling answers. According to this model, if people do not possess the necessary knowledge to produce a normatively correct response, their errors derive from a *mindware gap* (i.e. missing knowledge). Referring to the previous example, the intuitive answer might be the only one when the fundamental rule of proportions is missing. When relevant knowledge and procedures are not available (i.e. they are not learned or poorly compiled), we cannot have an override since to override the intuitive response a different response is needed as a substitute.

However, even if people detect the conflict between their intuitions and a normative rule, and thus relevant knowledge is available, they can still produce a normatively incorrect response. In this case, their errors result from an *override failure*: Different alternatives are produced and there is an attempt to override Type 1 processing, but this attempt fails usually because people do not have the necessary cognitive capacity to inhibit beliefs, feelings, and impressions, and at the same time, to implement the appropriate normative rules (e.g. De Neys et al. 2005; Handley et al. 2004). So, we have an override failure when people hold the rule but they do not base their answer on it.

Stanovich and colleagues (2008) outlined how people can reach a correct solution when they have necessary cognitive capacity to inhibit competing responses and to use their mindware to solve the task. A number of studies (see Stanovich and West 2000 for a review) found evidence that people with higher cognitive capacity will be more likely to produce normatively correct responses. Kahneman and Frederick (2002) pointed out that higher ability people are more likely to possess the relevant logical rules and also to recognise the applicability of these rules (i.e. they are more likely to overcome erroneous intuitions). Therefore, thinking errors are expected to decrease with increasing cognitive ability (Evans et al. 2009; Morsanyi and Handley

2008). According to Stanovich and West (2008) because it requires considerable cognitive resources carrying out slow, sequential and effortful Type 2 computations, while simultaneously inhibiting quick, low-effort, and intuitively compelling Type 1 responses.

Finally, sometimes errors arise from the use of inappropriate knowledge and strategies that people hold and drive reasoning processes far from the logical standpoint. In this case, the failure is related to *contaminated mindware* in Stanovich et al.'s model. There are various mechanisms that can lead to "contamination". Toplak et al. (2007) propose that a good candidate for contaminated mindware in the case of probabilistic reasoning could be superstitious thinking. In the previous example, the urn with more red marbles might be chosen because the respondent holds the superstitious belief that red is a lucky colour.

Starting from these premises, we present a set of experiments designed to test the Stanovich and colleagues' model inside probabilistic reasoning. The model of Stanovich and colleagues provides a theoretical framework for reconciling the educational and dual-process approaches. Indeed, studies that explore the impact of education on probabilistic reasoning (e.g. Fischbein and Schnarch 1997; Lehman et al. 1988) usually do not investigate the interactions between level of education, cognitive capacity and thinking styles. By contrast, studies inside the dual approach framework typically focus on the effect of cognitive ability, cognitive load, and thinking styles on adults' reasoning including sometimes age-related changes (e.g. Jacobs and Klaczynski 2002; Klaczynski 2009; Brainerd and Reyna 2001; Reyna and Farley 2006).

However, although Stanovich and colleagues (2008) offer a useful framework for investigating the interplay between these factors, they do not make specific predictions regarding changes based on the educational level. Thus, one important aim of the present series of experiments is to investigate knowledge, cognitive capacity and thinking styles simultaneously in population characterized by different educational levels. In particular, we investigated the *mindware gap* (i.e. the missing knowledge) taking into account individual differences in cognitive ability and *contaminated mindware* (i.e. the superstitious beliefs) (Study 1). Then, we explored the *override failure* stressing the role of mindware and the need to use it (Study 2).

1 Study 1: The Mindware Gap

Since mathematical abilities (that is, the mindware) concerned with probabilistic reasoning are learned and consolidated through education, we carried on the research with students of different grade levels. Thus, in the present study we considered children's grade level as an indicator of their knowledge regarding probability. Three experiments were conducted with primary, secondary, and high school students that were presented age-adapted probabilistic reasoning tasks. We expected that younger students, whose computational capacities involved in probabilistic reasoning are less consolidated, should perform worse than older ones. Their performance should be explained respectively by *mindware gap* and mindware availability.

In order to better ascertain the role of mindware, we controlled the impact of cognitive ability and superstitious thinking. In fact, cognitive ability should have a positive effect, whereas superstitious thinking should have a detrimental effect on probabilistic reasoning performance (see Toplak et al. 2007).

1.1 Experiment 1

1.1.1 Method

Participants The participants were 241 primary school students enrolled in primary schools that serve families from lower middle to middle socioeconomic classes in Tuscany, Italy. Children attended grade 3 ($N = 133$, 51 % boys; mean age: 8.3 yrs, $SD = 0.56$) and grade 5 ($N = 118$, 55 % boys; mean age: 10.5 yrs, $SD = 0.50$). These grade levels were chosen since some basics of probability are taught to the fourth and fifth graders following the Italian national curricular programs.[1] Then, we included in the sample students before they were taught probability issues (third graders), and students who had been taught probability issues (fifth graders). Children's parents were given information about the study and their permission was requested.

Measures *Gambler Fallacy Task* (Primi and Chiesi 2011). Following several studies that have measures gambler fallacy in children (e.g. Afantiti-Lamprianou and Williams 2003; Batanero et al. 1994) and college students (Konold 1995), we developed a specific task. A preliminary version of this task was used in a previous study run with children and college students (Chiesi and Primi 2009). It consists in a marble bag game in which different base-rates in combination with two different sequences of outcomes were used. In detail, it was composed of 3 different trials in which the proportion of Blue (B) and Green (G) marbles varied (15B & 15G; 10B & 20G; 25B & 5G). Thus, the present task allows for testing the gambler fallacy with both equally likely and not equally likely proportions. Before the task was presented, children were shown a video in which the marble bag game was played. The bag shown in the video has a see-through corner and instead of drawing a marble from the bag, the marble is pushed into that corner and then moved back. Since the bag remains always closed, visibly the number of the marbles remains always the same. After the video, each participant received a sheet where it was written the following instruction: "*15 blue and 15 green marbles have been put into the bag shown*

[1]Specifically, the curriculum include statistical surveys and their representations, some linguistic/conceptual issues related to possible, impossible, improbable events, and the development of judgement under uncertainty and estimation of odds through games of chance, inside the classical definition of probability.

in the video and one ball has been pushed in the see-through part. It was done a few times and a sequence of 5 green marbles was obtained". The question was: "*The next one is more likely to be…*". The following instruction explain that: "*The game was repeated again and a sequence of 5 blue marbles was obtained*". Then the question was: "*The next one is more likely to be…*". After this first trial, the two other trials were presented changing the proportion of blue and green marbles. For each trial the same questions of the first trial were asked. We formed a composite scores (range 0–6) summing correct answers that represent normative reasoning, i.e. the higher the score, the higher the respondent's ability to avoid the gambler fallacy.

Set I of the Advanced Progressive Matrices (APM-Set I, Raven 1962). To measure children's cognitive abilities the APM-Set I was administered as a short form of the Raven's Standard Progressive Matrices (SPM, Raven 1941; for a detailed analysis to test its suitability as short form see Chiesi et al. 2012b). The Set I of APM is composed of 12 matrices increasing in their difficulty level, and the items covered the range of difficulty of SPM (Raven 1962). These items are composed of a series of perceptual analytic reasoning problems, each in the form of a matrix. The problems involve both horizontal and vertical transformation: figures may increase or decrease in size, and elements may be added or subtracted, flipped, rotated, or show other progressive changes in the pattern. In each case, the lower right corner of the matrix is missing and the participant's task is to determine which of eight possible alternatives fits into the missing space such that row and column rules are satisfied. A score ranging from 0 to 12 was obtained summing the correct answers.

Superstitious Thinking Scale (Kokis et al. 2002). The scale was composed of 8 items referring to superstitious beliefs and luck. Example items are: "I have things that bring me luck" (positively scored), "I do not believe in luck" (negatively scored). The Italian version of the scale was obtained using a forward-translation method and validated through a sample of students from third to eighth grade (Chiesi et al. 2010).

Procedure The Gambler Fallacy task was presented first and then superstitious thinking and cognitive ability were measured using the above described scales. Participants completed them in a single session during school time. The session took about 30–35 minutes altogether. Tasks were collectively administered and presented in a paper and pencil version, and students had to work through them individually.

1.1.2 Results and Discussion

Correlations between the variables measuring probabilistic reasoning and individual differences in cognitive ability and superstitious thinking were computed. Probabilistic reasoning was not correlated with superstitious thinking ($r(N = 249) = -0.11, n.s.$), and it was positively correlated ($r(N = 249) = 0.41$, $p < 0.001$) with cognitive ability. Then, to examine the effect of grade levels (3 and 5) on probabilistic reasoning, a one-way ANCOVA was run in which only cognitive ability was used as covariate. The effect of grade was significant ($F(1, 248) = 6.53$, $p < 0.05$,

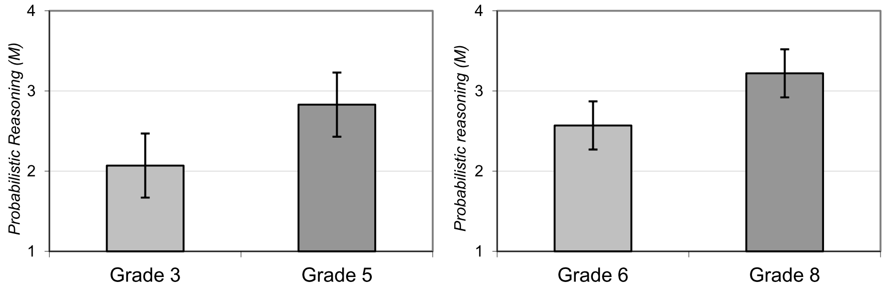


Fig. 2 Mean values representing probability reasoning for each grade group in Experiment 1 (*left*) and Experiment 2 (*right*). Standard errors are represented in the figure by the *error bars* attached to each column

$\eta_p^2 = 0.03$) once the significant effect of cognitive ability ($F(1, 248) = 31.24$, $p < 0.001$, $\eta_p^2 = 0.11$) was partialled out. As expected, grade 5 showed a better performance in probabilistic reasoning ($M = 2.83$, $SD = 1.52$) than grade 3 ($M = 2.07$, $SD = 0.87$). That is, after taking into account the effect of cognitive ability, the grade level continued to have a significant effect on children's reasoning performance (see Fig. 2).

As in previous studies (e.g. Handley et al. 2004; Kokis et al. 2002; Stanovich and West 1999), the present results confirmed the tendency of analytic processing to increase with increasing cognitive ability. Thus, our findings, in line with the argument of Kahneman and Frederick (2002), showed that children with higher cognitive ability are more likely to possess the relevant mathematical and probabilistic rules, as well as to recognise the applicability of these rules in particular situations. Additionally, our results point out that changes in probabilistic reasoning ability are related to new acquired and consolidate mindware from lower to higher grades. Finally, superstitious thinking (see Toplak et al. 2007) did not act as *contaminated mindware*. Arguably, at this age children's judgements were not affected by impression or feelings related to luck or false beliefs about random events.

1.2 Experiment 2

1.2.1 Method

Participants The experiment was conducted with a sample of secondary school students enrolled in schools that serve families from lower middle to middle socioeconomic classes in the same area of the Experiment 1. Students attended grade 6 ($N = 82$, 52 % boys; mean age: 11.8 yrs, $SD = 0.51$) and grade 8 ($N = 121$, 56 % boys; mean age: 13.7 yrs, $SD = 0.57$). These grade levels were chosen since from the 6 to 8 grade Italian students consolidate the use of fractions and proportions that represent the prerequisites for probabilistic reasoning. Students' parents were given information about the study and their permission was requested.

Table 1 Summary of the tasks employed in Experiment 2 and 3 with reference to the origin, requested rule, and related biases

	Source	Rule	Bias
Task 1	Kahneman et al. (1982)	likelihood of independent and equiprobable events	*gambler's fallacy*
Task 2	Kahneman et al. (1982)	likelihood of strings of independent and equiprobable events	*random similarity bias*
Task 3	Denes-Raj and Epstein (1994)	ratios computation and comparison:	*ratio bias*
Task 4	Kahneman and Tversky (1973)	likelihood of one event	*base-rate fallacy*
Task 5	Tversky and Kahneman (1983)	conjunction rule	*conjunction fallacy*
Task 6	Green (1982)	likelihood of one event	*equiprobability bias*

Measures and Procedure Different tasks were employed in order to measure probabilistic reasoning. Tasks were collectively administered and presented in a paper and pencil version, and children had to work through them individually. The students worked through 6 different probabilistic reasoning tasks (Chiesi et al. 2011) adapted from the heuristics and biases literature, as the gambler fallacy or the base-rate fallacy (see Table 1 for the source of each task, the normative principles required to solve them), and the related biases. There were three response options in the case of each task, and children were given either 1 (correct) or 0 (incorrect) points for each response. As in previous studies (Kokis et al. 2002; Toplak et al. 2007; West et al. 2008) the score on the six probabilistic reasoning tasks were summed to form a composite score (range 0–6).

The scales to measure superstitious thinking and cognitive ability as well as the procedure are described in Experiment 1. The administration time for this experiment was about 35–40 minutes.

1.2.2 Results and Discussion

Probabilistic reasoning was negatively correlated with superstitious thinking ($r(N = 193) = -0.27$, $p < 0.001$) and positively correlated with cognitive ability ($r(N = 193) = 0.34$, $p < 0.001$). Then, to examine the effect of grade levels (6 and 8) on probabilistic reasoning, a one-way ANCOVA was run in which both cognitive ability and superstitious thinking were used as covariates. The effect of grade was significant ($F(1, 192) = 7.81$, $p < 0.01$, $\eta_p^2 = 0.04$) once the significant effects of cognitive ability ($F(1, 192) = 12.17$, $p < 0.001$, $\eta_p^2 = 0.06$) and superstitious thinking ($F(1, 192) = 6.93$, $p < 0.01$, $\eta_p^2 = 0.03$) were partialled out. As we expected, higher grade children were more competent ($M = 3.22$, $SD = 1.20$) than lower grade children ($M = 2.57$, $SD = 1.22$) in solving probabilistic reasoning tasks (Fig. 2).

The present results showed, in line with previous studies (e.g. Handley et al. 2004; Kokis et al. 2002; Stanovich and West 1999) and Experiment 1, that children with higher cognitive ability performed better and, differently from Experiment 1, superstitious thinking deteriorated probabilistic reasoning performance, playing a role as *contaminated mindware*. However, the main effect of grade level indicated that some rules need to be taught and exerted.

In sum, Experiment 1 and Experiment 2 demonstrated an effect of education on probabilistic reasoning once controlled the effect of individual differences in cognitive ability and superstitious thinking.

1.3 Experiment 3

In the previous experiments, in order to investigate *mindware gap*, we used the grade level as an indirect measure of the acquisition and consolidation of knowledge related to probability. In this experiment, to establish if people lack or hold the mindware, we measured directly the relevant knowledge needed for dealing with the probability tasks they were asked to solve.

1.3.1 Method

Participants The experiment was conducted with a sample of high school students ($N = 372$, 68 % boys; mean age: 16.3 yrs, $SD = 0.89$) enrolled in schools that serve families from lower middle to middle socioeconomic classes in the same area of the previous experiments. We employed a sample of high school students in order to work with students who had encountered issues related to probability throughout primary to high school years. Parents of minors and students aged higher than 18 years were given information about the study and required a consent form.

Measures and Procedure To measure probabilistic reasoning we employed the *Gambler Fallacy Task* described in Experiment 1 and the tasks described in Experiment 2. We obtained a composite score summing the correct answers (range 0–12). Before performing the tasks, participants were presented four questions measuring knowledge of basic mathematical principles involved in probabilistic reasoning (i.e. the ability to reason correctly with proportions and percentages). An example of item was *"Smokers are 35 % of the population. There are 200 passengers on a train. How many of them will be smokers?"*. Students were given one point for each correct answer, thus a total score (ranged from 0 to 4) was obtained. This score was intended to measure the mindware.

Superstitious thinking was measured using the scale described in Experiment 1. To measure cognitive ability we employed the Advanced Progressive Matrices—Short Form (APM-SF; Arthur and Day 1994; for a detailed analysis of its suitability as short form see Chiesi et al. 2012a). The APM-SF consists of 12 items selected

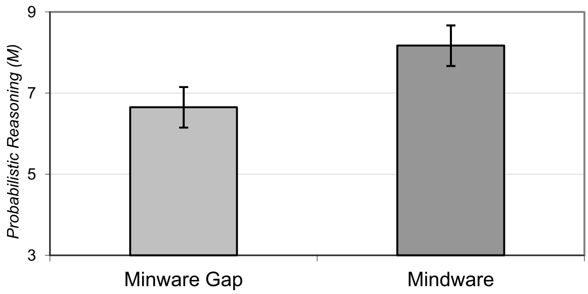


Fig. 3 Mean values representing probability reasoning for each group of high school students differentiated on mindware. Standard errors are represented in the figure by the *error bars* attached to each column

from the 36 items of the APM-Set II (Raven 1962). Participants have to choose the correct response out of eight possible options. A score ranging from 0 to 12 was obtained summing the correct answers.

For the procedure see Experiment 1. The administration session took about 40–45 minutes altogether.

1.3.2 Results and Discussion

Probabilistic reasoning was negatively correlated with superstitious thinking ($r(N = 363) = -0.20$, $p < 0.001$) and positively correlated with cognitive ability ($r(N = 363) = 0.32$, $p < 0.001$). Referring to the mindware scores, we created two groups: mindware gap (scores ranging from 0 to 3) and mindware (score equal to 4). To examine the effect of mindware gap and mindware on probabilistic reasoning, a one-way ANCOVA was run in which cognitive ability and superstitious thinking were used as covariates. The effect of mindware was still significant ($F(1, 354) = 19.16$, $p < 0.001$, $\eta_p^2 = 0.05$) once the significant effects of cognitive ability ($F(1, 354) = 28.71$, $p < 0.001$, $\eta_p^2 = 0.08$) and superstitious thinking ($F(1, 354) = 7.01$, $p < 0.01$, $\eta_p^2 = 0.02$) were partialled out. As we expected, the mindware group obtained higher scores ($M = 8.17$, $SD = 2.05$) than the mindware gap group ($M = 6.65$, $SD = 2.42$) in solving probabilistic reasoning tasks (Fig. 3).

In sum, Experiment 3 confirmed that individual differences in cognitive ability and superstitious thinking affected reasoning but that acquired rules (mindware) represent a necessary tool to deal with probability.

2 Study 2: The Override Failure

The findings of the previous study provide evidence for the relevance of the mindware in probabilistic reasoning. Nonetheless, from these results it is not possible to ascertain if some students were wrong because of override failures, that is, if some students hold the relevant mindware but they do not base their judgements on it. In fact, there is some evidence that even those who give incorrect responses experience a conflict between intuitions (related to Type 1 process) and logic (related to Type 2 process) (e.g. De Neys et al. 2008). That is, those who end up giving an incorrect response might spend some time evaluating the different response options, equally available, but eventually they chose the intuitively compelling one. Thus, an incorrect response does not necessarily imply that the normative option is missed but that it is just ignored, i.e. people may possess more knowledge about rules of probabilities than their answers show. The following three experiments aimed at better exploring this point.

One possible way of distinguishing between reasoning errors that arise from a lack of relevant knowledge and those that are the result of participants' not investing enough effort into implementing the rule properly is to use different instructional conditions. Dual-process theories predict that increasing cognitive effort (that is, increasing the amount of Type 2 processing) should lead to an increase in normative responding. In a study conducted with college students, Ferreira et al. (2006) found that instructions to be intuitive vs. rational oriented the tendency for using heuristic vs. rule-based reasoning when participants solved base-rate, conjunction fallacy, and ratio bias tasks. Similarly, Klaczynski (2001) reported that *framing* instructions to reason "like a perfectly logical person" boosted the performance of adolescents and adults on ratio bias problems.

The aim of the Experiment 4 and Experiment 5 was to ascertain if older secondary school children (i.e. those who possess more relevant knowledge) and college students (i.e. those who had acquired and exerted the relevant knowledge) were able to generate two responses, and to operate an *instruction oriented* choice asking them to answer in an intuitive or logical manner. Moreover, we aimed to verify if cognitive ability mediated the impact of instructions, and we explored these aspects controlling the effect of superstitious thinking. Finally, in Experiment 6 we assessed the effect of a training in prompting the override of the intuitive answer.

2.1 Experiment 4

2.1.1 Method

Participants The experiment was conducted with 126 secondary school students in grade 8 (48 % boys; mean age: 13.7 yrs, $SD = 0.59$) enrolled in schools that serve families from lower middle to middle socioeconomic classes in the same area of the previous experiments. We employed a sample of older high school students in order to work with students who were supposed to hold the relevant knowledge. All students' parents were given information about the study and their permission was requested.

Measures and Procedure We administered the same scales employed in Experiment 2 and 3 but, in order to obtain a more itemed measure of probabilistic reasoning, we added four tasks (Table 2) to the previous ones. The score on the ten probabilistic reasoning tasks were summed to form a composite score (range 0–10).

Table 2 Summary of the tasks added to those employed in Study 1 with reference to the origin, requested rule, and related biases

	Source	Rule	Bias
Task 7	Tversky and Kahneman (1974)	departures from population are more likely in small samples	*sample size neglect*
Task 8	Stanovich and West (2003)	likelihood of independent events	*random similarity bias*
Task 9	Kirkpatrick and Epstein (1992)	ratios computation and comparison	*ratio bias*
Task 10	Konold (1989)	likelihood of compound events	*equiprobability bias*

The procedure was the same as in Experiment 2 except that participants were given one of two different instructions (based on Klaczynski 2001; and Ferreira et al. 2006). In the intuitive condition, participants were told: *"Please answer the questions on the basis of your intuition and personal sensitivity."* In the rational condition, participants were told *"Please answer the questions taking the perspective of a perfectly logical and rational person."* As in Ferreira et al. (2006), a between-subjects design was used where participants were randomly assigned to the intuitive or to the rational condition. Sixty-three students were given rational instructions, and 63 students were given intuitive instructions.

2.1.2 Results and Discussion

In order to investigate the role of cognitive ability on the capacity to cope with the instruction, two groups were created by using the median (9) of the Raven's Matrices score as a cut-off: students below the median formed the low to medium cognitive ability group ($N = 54$, $M = 6.65$, $SD = 1.54$), and students at or above the median formed the high cognitive ability group ($N = 71$, $M = 10.19$, $SD = 0.96$).

To examine the effect of instructions (intuitive vs. rational) and cognitive ability (low vs. high) on probabilistic reasoning, a 2 × 2 ANCOVA was run in which superstitious thinking was used as a covariate. The effect of superstitious thinking was significant ($F(1, 124) = 3.96$, $p < 0.05$, $\eta_p^2 = 0.03$). Once this significant effect was partialled out, the main effect of instructions was not significant as well as the main effect of cognitive ability. The interaction between cognitive ability and instruction was significant ($F(1, 124) = 5.71$, $p < 0.05$, $\eta_p^2 = 0.05$). This was because higher ability children benefited from the instruction to reason logically (rational: $M = 5.17$, $SD = 1.36$, intuitive: $M = 4.43$, $SD = 1.32$) whereas in lower ability children there was no difference between the two instruction conditions (Rational: $M = 4.57$, $SD = 1.73$, Intuitive: $M = 4.13$, $SD = 1.20$). That is, whether students were asked to be rational or intuitive, they performed in the same way (Fig. 4).

The main aim of the present experiment was to manipulate the mental effort that participants invest in solving the tasks. The impact of instructions was moderated by cognitive ability (i.e. we did not find a main effect of instructions). Specifically, only

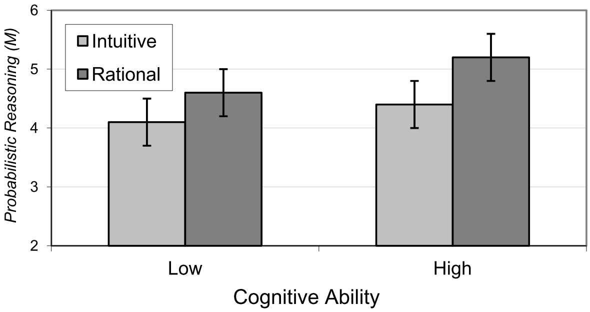


Fig. 4 Mean values representing probability reasoning for each cognitive ability groups (secondary school students). In each group, respondents were differentiated on the received instruction (intuitive or rational). Standard errors are represented in the figure by the *error bars* attached to each column

students with high cognitive capacity benefited from investing effort into thinking "like a perfectly logical person." The most straightforward explanation for this is that higher ability people are more likely to possess the relevant rules, and for them it is easier to recognise when these rules have to be implemented.

2.2 *Experiment 5*

2.2.1 Method

Participants The experiment was conducted with 60 college students (49 % men; mean age: 24.5 yrs, $SD = 3.70$) enrolled in different degree programs (Psychology, Educational Sciences, Biology, and Engineering) at the University of Florence. They were given information about the study and required to fill a consent form. All participants were volunteers and they did not receive any reward for their participation in this study.

Measures and Procedure The same ten probabilistic reasoning tasks used in Experiment 4 were administered. To measure cognitive ability we employed the Advanced Progressive Matrices—Short Form (APM-SF; Arthur and Day 1994) described in Experiment 3. Finally, students were administered the same scale used for children and adolescents to measure superstitious thinking (Chiesi et al. 2010) as done in previous experiments (Morsanyi et al. 2009).

The procedure was the same as in Experiment 4. Thirty students were given rational instructions, and 30 students were given intuitive instructions.

2.2.2 Results and Discussion

To investigate the role of cognitive ability on the instruction, two groups were created by using the median (9) of the Raven's Matrices score as a cut-off. Students below the median formed the low to medium cognitive ability group ($N = 29$, $M = 5.17$, $SD = 1.87$), and students at or above the median formed the high cognitive ability group ($N = 31$, $M = 10.16$, $SD = 1.39$).

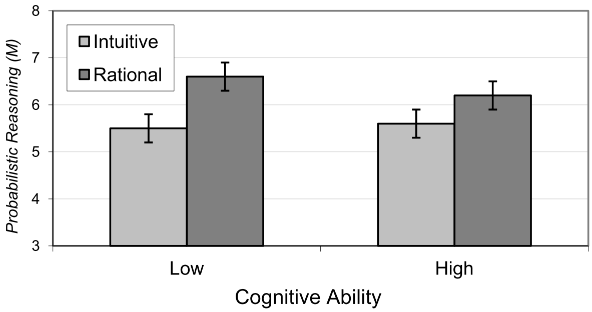


Fig. 5 Mean values representing probability reasoning for each cognitive ability group (college students). In each group, respondents were differentiated on the received instruction (intuitive or rational). Standard errors are represented in the figure by the *error bars* attached to each column

To examine the effect of instructions (intuitive vs. rational) and cognitive ability (low vs. high) on probabilistic reasoning, a 2 × 2 ANCOVA was run in which superstitious thinking was used as a covariate. The results indicated that the effect of superstitious thinking was not significant. There was also no effect of cognitive ability, whereas the main effect of instructions was significant ($F(1,58) = 4.64$, $p < 0.05$, $\eta_p^2 = 0.09$, rational: $M = 6.33$, $SD = 1.27$, intuitive: $M = 5.58$, $SD = 1.23$). No interaction between cognitive ability and instructions was found. That is, college students showed better performance when instructed to be rational (Fig. 5).

Based on the results of Experiment 4, we expected that, given their more consolidated knowledge of probability rules, college students would generally be able to follow instructions, that is, it easier for them to recognise when to apply the relevant normative rules. The results supported these predictions. Whereas, differently from Experiment 4, college students' performance depended mainly on instruction conditions, and there was no effect of superstitious thinking and cognitive ability. Indeed, college students were able to give more normative responses when instructed to reason logically, regardless their cognitive ability. Presumably, because they all had the necessary ability to follow the instructions.

In sum, Experiment 4 and Experiment 5 demonstrated that some students were wrong because of override failures. Thus, some students hold the relevant mindware but they do not base their judgements on it. In fact, when instructed, they were able to use their mindware and make a correct *instruction oriented* choice.

2.3 Experiment 6

In making override failures, people hold the relevant mindware but they are not able to override the compelling intuitive response since some judgement seem to be right regardless logical and rule-based considerations. Using different instructional conditions is one possible way of distinguishing between reasoning errors that arise from a lack of relevant knowledge and those that are the result of participants' not implementing the rule properly. Nonetheless, participants' potential interpretation of the problems and the experimental instructions might represent a limit of this procedure. Thus, instead of giving the instruction to reason logically, in the present experiment we aimed to assess the effect of a training in which the logical vs. intuitive solutions of the same task were presented and compared. In this way, we expected to prompt the override of the intuitive answer when dealing with random events. Given the effect of superstition, we also tried to reduce its impact on reasoning explaining the irrationality of some belief about random events.

2.3.1 Method

Participants Participants were 83 high school students (65 % boys, mean age: 16.1; $SD = 0.56$) enrolled in schools that serve families from lower middle to middle socioeconomic classes in the same area of the previous experiments. As stated above, we employed a sample of high school students in order to work with students who had encountered issues related to probability throughout primary to high school years. Parents of minors and students aged higher than 18 years were given information about the study and required to fill a consent form.

Measures and Procedure As in Experiment 3, participants were presented the four questions measuring knowledge of basic mathematical principles involved in probabilistic reasoning (mindware), and the superstitious thinking and cognitive ability scales. Participants completed this battery of questionnaires in a single session during school time. The session took about 25 minutes altogether.

Then, participants were randomly divided in two groups: the training group ($N = 36$) and the control group ($N = 48$). The training consisted in two units (one hour each). In the first unit, students made experiments with random generators (i.e. throwing dice, sorting a card from a deck) and they were invited to compare intuitions and rule-based considerations. In the second unit, they were showed the irrationality of the superstitious beliefs about random events providing evidences that the outcome of chance events cannot be influenced or controlled. The control group followed a lesson about risk and health behaviours in adolescence. After, in both groups, probabilistic reasoning was measured employing the *Gambler Fallacy Task* described in Experiment 1.

2.3.2 Results and Discussion

As a preliminary step, we had verified that there were not differences between training and control group as regards cognitive ability ($t(81) = 1.82$, *n.s.*), superstitious thinking ($t(81) = 1.16$, *n.s.*) and mindware ($t(81) = 0.90$, *n.s.*). After the experimental manipulation, correlations between the variables in the study—probabilistic reasoning and individual differences in cognitive ability and superstitious thinking—were computed separately for each group. In the training group, probabilistic reasoning was not correlated with superstitious thinking ($r(N = 36) = 0.10$, *n.s.*), whereas it was positively correlated with cognitive ability ($r(N = 36) = 0.47$, $p < 0.01$). In the control group, probabilistic reasoning was

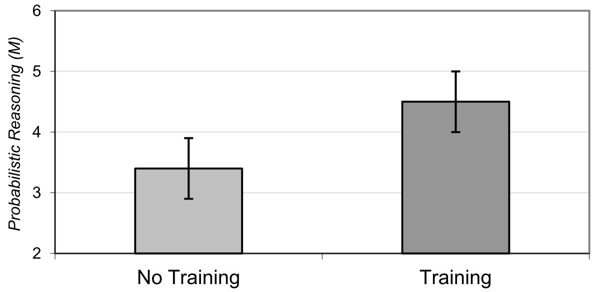


Fig. 6 Mean values representing probability reasoning for the group who followed the training and the control group (No training). Standard errors are represented in the figure by the *error bars* attached to each column

negatively correlated with superstitious thinking ($r(N = 47) = -0.29$, $p < 0.05$) and it was positively correlated with cognitive ability ($r(N = 47) = 0.30$, $p < 0.05$).

To examine the effect of training on probabilistic reasoning, a one-way ANCOVA was run in which superstitious thinking and cognitive ability were used as covariates. The effect of training was significant ($F(1, 78) = 4.31$, $p < 0.05$, $\eta_p^2 = 0.05$) once the significant effect of cognitive ability ($F(1, 78) = 11.76$, $p < 0.001$, $\eta_p^2 = 0.14$) was partialled out (the effect of the other covariate, i.e. superstitious thinking, was not significant). As we expected, the training group obtained higher scores ($M = 4.50$, $SD = 1.99$) than the control group ($M = 3.37$, $SD = 1.81$) in solving the probabilistic reasoning task (Fig. 6).

As expected, activities in which the logical and the intuitive approach to randomness were experienced prompted the override of the intuitive answer. Looking at correlations in each group, this result can be also referred to a possible effect of training in preventing the detrimental effect of superstitious beliefs on probabilistic reasoning.

3 General Discussion

The model of Stanovich and colleagues provides a theoretical framework for reconciling the educational and dual-process approaches. In a series of experiments, we examined the interactions between level of education, cognitive capacity and superstitious thinking in determining reasoning performance and the related errors. As detailed below, the results of the experiments presented in Study 1 (referring to the *mindware gap*) and Study 2 (referring to the *override failure*) are consistent with this model.

Referring to Study 1, in Experiment 1 and Experiment 2, we investigated the effect of grade levels on probabilistic reasoning ability, while controlling for the effects of cognitive capacity and superstitious thinking. Using different age-adapted tasks and scales, we found that grade levels accounted for a significant proportion of variance in probabilistic reasoning once the significant effects of cognitive ability and superstitious thinking were removed. The effect of grade levels might be explained referring to the increases of probability knowledge with education, and supports the assumption that relevant mindware plays an important role in probabilistic reasoning. This finding was confirmed in Experiment 3 in which the relevant mindware was directly measured. Additionally, our findings are in line with the claim that cognitive capacity will be good a predictor of normative reasoning only if the relevant knowledge to solve a task is acquired (see Stanovich and West 2008).

Referring to Study 2, in Experiments 4 and 5, we manipulated experimentally the effort that respondents invested in solving the problems by providing them with instructions to reason rationally/intuitively. Instructions to reason rationally have been found to increase normative performance as in Ferreira et al. (2006) and Klaczynski (2001), especially in the case of higher ability participants, as in Morsanyi et al. (2009) and Chiesi et al. (2011). In the terms of Stanovich and colleagues' (2008) model, increasing the mental effort that participants invest into reasoning will reduce the *override failures*, i.e. people succeed in overriding Type 1 processing and at the same time are able to implement the appropriate normative rules. In Experiment 6, we confirmed that performance was boosted when cues to resist tempting heuristic responses were given, and we confirmed that cognitive capacity will be a good predictor of normative reasoning only when the relevant mindware is available.

Overall the current studies, according to Stanovich et al. (2008), suggest that the correct solution can be reached holding the relevant mindware and recognising the need to use it and that individual differences in thinking styles and cognitive ability can be accounted for explaining some thinking errors. Additionally, they lend support to the claim of developmental dual-process theorists (e.g. Brainerd and Reyna 2001; Klaczynski 2009) that cognitive capacity *per se* is insufficient to explain changes in reasoning performance, because normative responding crucially depends on participants' relevant knowledge.

Although dual-process theories are very popular in many different areas of psychological science, including social, developmental and cognitive psychology, these theories are not without controversy (see, e.g. Keren and Schul 2009; Osman and Stavy 2006), as well as the use of the classic problems in the heuristics and biases tradition that has been criticized on the basis that they create an unnatural conflict between pragmatic/interpretative processes (see, e.g. Hertwig and Gigerenzer 1999). Nonetheless, we think that conceptualising probabilistic reasoning as an interplay between intuitive and rule-based processes offers some cues to cross the bridge from a psychological approach to an educational approach. From a psychological standpoint, the relevance of mindware in probabilistic reasoning stresses the role of education since probability rules might be very hard to derive from the experience. Thus, what learned at school about normative probabilistic reasoning assume a relevance in everyday life in which the ability to make decisions on the basis of probabilistic information is extremely important and the inability to make optimal choices can be extremely costly. From an educational perspective, psychology helps in understanding why probability is a hard subject to learn and teach (e.g. Kapadia and Borovcnik 1991; Shaughnessy 1992). Thus, teachers have to know that students are *naturally* inclined to rely on intuitions, and that normative rules are often at odds with these intuitions that appear to be right regardless rule-based considerations.

References

Afantiti-Lamprianou, T., & Williams, J. (2003). A scale for assessing probabilistic thinking and the representativeness tendency. *Research in Mathematics Education*, *5*, 172–196.

Arthur, W., & Day, D. (1994). Development of a short form for the Raven Advanced Progressive Matrices test. *Educational and Psychological Measurement*, *54*, 395–403.

Batanero, C., Godino, J. D., Vallecillos, A., Green, D. R., & Holmes, P. (1994). Errors and difficulties in understanding elementary statistical concepts. *International Journal of Mathematical Education in Science and Technology*, *25*, 527–547.

Brainerd, C. J., & Reyna, V. F. (2001). Fuzzy-trace theory: dual processes in memory, reasoning, and cognitive neuroscience. *Advances in Child Development and Behavior*, *28*, 49–100.

Chiesi, F., & Primi, C. (2009). Recency effects in primary-age children and college students using a gaming situation. *International Electronic Journal of Mathematics Education*, *4*, 3. Special issue on "Research and development in probability education". www.iejme.com.

Chiesi, F., Donati, M., Papi, C., & Primi, C. (2010). Misurare il pensiero superstizioso nei bambini: validità e attendibilità della Superstitious Thinking Scale (Measuring superstitious thinking in children: validity and reliability of the superstitious thinking scale). *Età Evolutiva*, *97*, 5–14.

Chiesi, F., Primi, C., & Morsanyi, K. (2011). Developmental changes in probabilistic reasoning: the role of cognitive capacity, instructions, thinking styles and relevant knowledge. *Thinking & Reasoning*, *17*, 315–350.

Chiesi, F., Ciancaleoni, M., Galli, S., Morsanyi, K., & Primi, C. (2012a). Item response theory analysis and differential item functioning across age, gender and country of a short form of the advanced progressive matrices. *Learning and Individual Differences*, *22*, 390–396. doi:10.1016/j.lindif.2011.12.007.

Chiesi, F., Ciancaleoni, M., Galli, S., & Primi, C. (2012b). Using the advanced progressive matrices (set I) to assess fluid ability in a short time frame: an item response theory-based analysis. *Psychological Assessment*, *24*, 892–900. doi:10.1037/a0027830.

Denes-Raj, V., & Epstein, S. (1994). Conflict between intuitive and rational processing: when people behave against their better judgment. *Journal of Personality and Social Psychology*, *66*, 819–829.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2005). Working memory and everyday conditional reasoning: retrieval and inhibition of stored counterexamples. *Thinking & Reasoning*, *11*, 349–381.

De Neys, W., Vartanian, O., & Goel, V. (2008). Smarter than we think: when our brains detect that we are biased. *Psychological Science*, *19*, 483–489.

Epstein, S. (1994). An integration of the cognitive and psychodynamic unconscious. *The American Psychologist*, *49*, 709–724.

Evans, J. St. B. T. (2006). The heuristic–analytic theory of reasoning: extension and evaluation. *Psychological Bulletin*, *13*(3), 378–395.

Evans, J. St. B. T. (2003). In two minds: dual process accounts of reasoning. *Trends in Cognitive Sciences*, *7*, 454–459.

Evans, J. St. B. T., & Over, D. E. (1996). *Rationality and reasoning*. Hove: Psychology Press.

Evans, J. St. B. T., Handley, S. J., Neilens, H., & Over, D. E. (2009). The influence of cognitive ability and instructional set on causal conditional inference. *Quarterly Journal of Experimental Psychology*, *63*, 892–909.

Ferreira, M. B., Garcia-Marques, L., Sherman, S. J., & Sherman, J. W. (2006). Automatic and controlled components of judgment and decision making. *Journal of Personality and Social Psychology*, *91*, 797–813.

Fischbein, E., & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for Research in Mathematics Education*, *28*(1), 96–105.

Green, D. R. (1982). *Probability concepts in 11–16 year old pupils* (2nd ed.). Loughborough: Centre for Advancement of Mathematical Education in Technology, College of Technology.

Handley, S., Capon, A., Beveridge, M., Dennis, I., & Evans, J. St. B. T. (2004). Working memory, inhibitory control and the development of children's reasoning. *Thinking & Reasoning*, *10*, 175–195.

Hertwig, R., & Gigerenzer, G. (1999). The "conjunction fallacy" revised: how intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making*, *12*, 275–305.

Jacobs, J. E., & Klaczynski, P. A. (2002). The development of decision making during childhood and adolescence. *Current Directions in Psychological Science*, *4*, 145–149.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: attribute substitution in intuitive judgement. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: the psychology of intuitive judgment* (pp. 49–81). New York: Cambridge University Press.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: heuristics and biases*. New York: Cambridge University Press.

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

Kapadia, R., & Borovcnik, M. (1991). *Chance encounters: probability in education*. Dordrecht: Kluwer Academic.

Keren, G., & Schul, Y. (2009). Two is not always better than one: a critical evaluation of two-system theories. *Perspectives on Psychological Science*, *4*, 533–550.

Kirkpatrick, L. A., & Epstein, S. (1992). Cognitive-experiential self-theory and subjective probability: further evidence for two conceptual systems. *Journal of Personality and Social Psychology*, *63*, 534–544.

Klaczynski, P. A. (2001). Framing effects on adolescent task representations, analytic and heuristic processing, and decision making: implications for the normative/descriptive gap. *Journal of Applied Developmental Psychology*, *22*, 289–309.

Klaczynski, P. A. (2004). A dual-process model of adolescent development: implications for decision making, reasoning, and identity. In R. V. Kail (Ed.), *Advances in child development and behavior* (Vol. 31, pp. 73–123). San Diego: Academic Press.

Klaczynski, P. A. (2009). Cognitive and social cognitive development: dual-process research and theory. In J. B. St. T. Evans & K. Frankish (Eds.), *In two minds: dual processing and beyond* (pp. 265–292). Oxford: Oxford University Press.

Kokis, J. V., MacPherson, R., Toplak, M. E., West, R. F., & Stanovich, K. E. (2002). Heuristic and analytic processing: age trends and associations with cognitive ability and cognitive styles. *Journal of Experimental Child Psychology*, *83*, 26–52.

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

Konold, C. (1995). Issues in Assessing Conceptual Understanding in Probability and Statistics. *Journal of Statistics Education*, *3*(1). Retrieved from http://www.amstat.org/publications/jse/v3n1/konold.html, May 25, 2009.

Lehman, D. R., Lempert, R. O., & Nisbett, R. E. (1988). The effects of graduate training on reasoning: formal discipline and thinking about everyday-life events. *The American Psychologist*, *43*, 431–442.

Morsanyi, K., & Handley, S. J. (2008). How smart do you need to be to get it wrong? The role of cognitive capacity in the development of heuristic-based judgment. *Journal of Experimental Child Psychology*, *99*, 18–36.

Morsanyi, K., Primi, C., Chiesi, F., & Handley, S. (2009). The effects and side-effects of statistics education. Psychology students' (mis-)conceptions of probability. *Contemporary Educational Psychology*, *34*, 210–220.

Osman, M., & Stavy, R. (2006). Intuitive rules: from formative to developed reasoning. *Psychonomic Bulletin & Review*, *13*, 935–953.

Perkins, D. N. (1995). *Outsmarting IQ: the emerging science of learnable intelligence*. New York: Free Press.

Primi, C., & Chiesi, F. (2011). The role of relevant knowledge and cognitive ability in gambler fallacy. In *The 7th congress of the European Society for Research in Mathematics Education*, Rzeszów, Poland. http://www.cerme7.univ.rzeszow.pl/WG/5/CERME_Primi-Chiesi.pdf.

Raven, J. C. (1941). Standardization of progressive matrices. *British Journal of Medical Psychology*, *19*, 137–150.

Raven, J. C. (1962). *Advanced progressive matrices*. London: Lewis & Co. Ltd.

Reyna, V. F., & Farley, F. (2006). Risk and rationality in adolescent decision-making: implications for theory, practice, and public policy. *Psychological Science in the Public Interest*, *7*, 1–44.

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

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah: Erlbaum.

Stanovich, K. E., & West, R. F. (1999). Discrepancies between normative and descriptive models of decision making and the understanding acceptance principle. *Cognitive Psychology*, *38*, 349–385.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: implications for the rationality debate. *Behavioral and Brain Sciences*, *23*, 645–726.

Stanovich, K. E., & West, R. F. (2003). Evolutionary versus instrumental goals: how evolutionary psychology misconceives human rationality. In D. E. Over (Ed.), *Evolution and the psychology of thinking* (pp. 171–230). Hove: Psychology Press.

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology*, *94*, 672–695.

Stanovich, K. E., Toplak, M. E., & West, R. F. (2008). The development of rational thought: a taxonomy of heuristics and biases. *Advances in Child Development and Behavior*, *36*, 251–285.

Thompson, V. A. (2009). Dual process theories: a metacognitive perspective. In J. Evans & K. Frankish (Eds.), *Two minds: dual processes and beyond*. London: Oxford University Press.

Toplak, M., Liu, E., Macpherson, R., Toneatto, T., & Stanovich, K. E. (2007). The reasoning skills and thinking dispositions of problem gamblers: a dual-process taxonomy. *Journal of Behavioral Decision Making*, *20*, 103–124.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, *185*, 1124–1131.

Tversky, A., & Kahneman, D. (1983). Extentional versus intuitive reasoning: the conjunction fallacy in probability judgment. *Psychological Bulletin*, *90*, 293–315.

West, R. F., Toplak, M. E., & Stanovich, K. E. (2008). Heuristics and biases as measures of critical thinking: associations with cognitive ability and thinking dispositions. *Journal of Educational Psychology*, *100*, 930–941.

Revisiting the Medical Diagnosis Problem: Reconciling Intuitive and Analytical Thinking

Lisser Rye Ejersbo and Uri Leron

Abstract A recurrent concern in mathematics education—both theory and practice—is a family of mathematical tasks which elicit from most people strong immediate ("intuitive") responses, which on further reflection turn out to clash with the normative analytical solution. We call such tasks *cognitive challenges* because they challenge cognitive psychologists to postulate mechanisms of the mind which could account for these phenomena. For the educational community, these cognitive challenges raise a corresponding *educational challenge*: What can we as mathematics educators do in the face of such cognitive challenges? In our view, pointing out the clash is not enough; we'd like to help students build bridges between the intuitive and analytical ways of seeing the problem, thus hopefully creating a peaceful co-existence between these two modes of thought. In this article, we investigate this question in the context of probability, with special focus on one case study—*the Medical Diagnosis Problem*—which figures prominently in the cognitive psychology research literature and in the so-called rationality debate. Our case study involves a combination of theory, design and experiment: Using the extensive psychological research as a theoretical base, we design a new "bridging" task, which is, on the one hand, formally equivalent to the given "difficult" task, but, on the other hand, is much more accessible to students' intuitions. Furthermore, this new task would serve as a "stepping stone", enabling students to solve the original difficult task without any further explicit instruction. These design requirements are operationalized and put to empirical test.

L.R. Ejersbo (✉)
DPU, Aarhus University, Copenhagen, Denmark
e-mail: LRE@dpu.dk

U. Leron
Israel Institute of Technology, Haifa, Israel
e-mail: uril@technion.ac.il

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_12,

1 Introduction: Cognitive and Educational Challenges

A recurrent concern in mathematics education—both theory and practice—is a family of mathematical tasks which elicit from most people strong immediate responses, which on further reflection, however, turn out to clash with the normative analytical solution. We will call such tasks *cognitive challenges*[1] because they challenge cognitive psychologists to postulate mechanisms of our mind which could account for these phenomena. Indeed, an extensive empirical and theoretical research program on cognitive challenges (or cognitive *biases*) has been going on in cognitive psychology since the second half of the last century, culminating in the 2002 Nobel prize in economy to Daniel Kahneman for his work with Amos Tversky on "intuitive judgment and choice" (Kahneman 2002). For the educational community, these cognitive challenges raise a corresponding *educational challenge*, which will be the focus of this article: What can we as mathematics educators do in the face of such cognitive challenges? We investigate this question via several examples, leading up to an in-depth probe of one case study—the Medical Diagnosis Problem (MDP)—which figures prominently in the cognitive psychology research literature.

The plan of the article is as follows. Section 2 contains the theoretical background material, beginning with two topics from the relevant research in cognitive psychology. In Sect. 2.1, we introduce the general theory concerning cognitive challenges (the so-called *dual-process theory*), and in Sect. 2.2, we zoom in on the medical diagnosis problem and survey the extensive literature surrounding it. Finally, in Sect. 2.3, we discuss some links with relevant research from the mathematics education community. Following the theoretical introduction in Sect. 2, the main part of the article (Sect. 3) is concerned with the corresponding *educational challenge*, which we now describe in more detail.

Suppose you have just presented one of these problems to your math class. Being a seasoned math teacher, you first let them be surprised; that is, you first let them do some guessing, bringing out the intuitive solution. Then you prompt them to do some reasoning and some calculations, leading to the analytical solution, which turns out to clash with the previous intuitive solution. Given the classroom situation just described, we are now faced with a crucial educational challenge: As teachers and math educators, what do we do next? We haven't conducted a survey, but it is our experience that most teachers would leave it at that, or at best discuss with the students the clash they have just experienced between the intuitive and analytical solutions. But is this the best we can do? Inspired by Papert's (1993/1980, pp. 146–150) ingenious treatment of the string-around-the-earth puzzle (see Sect. 3.1 below), we claim that we can actually do better. We want to avoid the "default" conclusion that students should not trust their intuition, and should abandon it in the face of conflicting analytical solution. We also want to help students overcome the uncomfortable situation, whereby their mind harbors two conflicting solutions, one intuitive but wrong, the other correct but counter-intuitive.

[1]This term comes from an ongoing work with Abraham Arcavi.

In addition to these pedagogical considerations, research shows that the rejected intuitions (often called misconceptions, pre-conceptions, or naïve ideas) are not eliminated, only go underground and will re-surface under unexpected conditions (see Sect. 2.3 below). Moreover, as we will show in the case of the MDP, these intuitions can serve as a stepping stone for building more advanced expert knowledge. This way, students' intuitions are viewed as a powerful resource, skills that students are good at and feel good about, rather than obstacle.

Indeed, we are interested in helping students build bridges between the intuitive and analytical ways of seeing the problem, thus hopefully creating a peaceful coexistence between these two modes of thought. As we will demonstrate in Sect. 3, our way of working towards such reconciliation involves a *design issue*: Using the extensive psychological research as a theoretical base, design a new "bridging" task, which is, on the one hand, formally equivalent to the given "difficult" task, but, on the other hand, is much more accessible to students' intuitions. Such tasks help mediate between the two modes of thinking about the problem and, furthermore, may help many students "stretch" their intuition towards a normatively correct solution. The design stage is followed by an experiment, which demonstrated that the new bridging task does satisfy the design criteria.

Since the MDP occupies a central part of the paper, it is important to clarify its status in this article. The article is definitely not *about* the MDP, nor is there any claim (at this preliminary stage) for generalizability—this would require a much more extensive research and development effort. Rather, the MDP serves here in the role of a case study, a generic example if you will, or—to use the language of engineers—*a proof of concept*.[2] Like all good case studies, we hope that through an in-depth analysis of one example we can gain some non-trivial insights, which can then be subjected to further research and generalizations.

2 Theoretical Background

2.1 Intuitive vs. Analytical Thinking: Dual-Process Theory (DPT)

To start off our theoretical discussion, consider the following puzzle:

> A baseball bat and ball cost together one dollar and 10 cents. The bat costs one dollar more than the ball. How much does the ball cost?

This simple arithmetical puzzle would be totally devoid of interest, if it were not for the *cognitive challenge* it presents, best summarized in Kahneman's (2002) Nobel Prize lecture:

---

[2]Wikipedia: A *proof of concept* [...] is realization of a certain method or idea(s) to demonstrate its feasibility, or a demonstration in principle, whose purpose is to verify that some concept or theory is probably capable of being useful. A proof-of-concept may or may not be complete, and is usually small and incomplete.

> Almost everyone reports an initial tendency to answer '10 cents' because the sum $1.10 separates naturally into $1 and 10 cents, and 10 cents is about the right magnitude. Frederick found that many intelligent people yield to this immediate impulse: 50 % (47/93) of Princeton students and 56 % (164/293) of students at the University of Michigan gave the wrong answer. (p. 451)

The trivial arithmetical challenge has thus turned into a non-trivial challenge for cognitive psychologists: What is it about the workings of our mind that causes so many intelligent people to err on such a simple problem, when they surely possess the necessary mathematical knowledge to solve it correctly?

Complicating this cognitive challenge even further, research in cognitive psychology has revealed that *harder* versions of the task may result in *better* performance by the subjects. For example, we can enhance the subjects' performance by making the numbers more messy (let the bat and ball cost together $1.12 and the bat cost 94 cents more than the ball), or by displaying the puzzle via hard-to-read font on a computer screen (Song and Schwarz 2008).

This challenge and many others like it have led to one of the most influential theories in current cognitive psychology, *Dual Process Theory* (DPT), roughly positing the existence of "two minds in one brain". These two thinking modes—intuitive and analytic—mostly work together to yield useful and adaptive behavior, but, as the long list of cognitive challenges demonstrate, they can also fail in their respective roles, yielding non-normative answers to mathematical, logical or statistical tasks. A corollary of particular interest for mathematics education is that many recurring and prevalent mathematical errors originate from general mechanisms of our mind and not from faulty mathematical knowledge. Significantly, such errors often result from the *strength* of our mind rather than its weakness (see the discussion of "good errors" in Gigerenzer 2005).

The distinction between intuitive and analytical modes of thinking is ancient, but has achieved a new level of specificity and rigor in what cognitive psychologists call *dual-process theory* (DPT). In fact, there are several such theories, but since the differences are not significant for the present discussion, we will ignore the nuances and will adopt the generic framework presented in Stanovich and West (2000), Kahneman and Frederick (2005) and Kahneman (2002). The following concise—and much oversimplified—introduction to DPT and its applications in mathematics education is abridged from Leron and Hazzan (2006, 2009). For state of the art thinking on DPT—history, empirical support, applications, criticism, adaptations, new developments—see Evans and Frankish (2009).

According to the dual-process theory, our cognition and behavior operate in parallel in two quite different modes, called *System 1* (S1) and *System 2* (S2), roughly corresponding to our commonsense notions of intuitive and analytical thinking. These modes operate in different ways, are activated by different parts of the brain, and have different evolutionary origins (S2 being evolutionarily more recent and, in fact, largely reflecting *cultural* evolution). The distinction between perception and cognition is ancient and well known, but the introduction of S1, which sits midway between perception and (analytical) cognition is relatively new, and has important consequences for how empirical findings in cognitive psychology are interpreted, including applications to the rationality debate (Samuels et al. 1999; Stanovich 2004) and to mathematics education research (Leron and Hazzan 2006, 2009).

Like perceptions, S1 processes are characterized as being fast, automatic, effortless, non-conscious and inflexible (hard to change or overcome); unlike perceptions, S1 processes can be language-mediated and relate to events not in the here-and-now. S2 processes are slow, conscious, effortful, computationally expensive (drawing heavily on working memory resources), and relatively flexible. In most situations, S1 and S2 work in concert to produce adaptive responses, but in some cases (such as the ones concocted in the heuristics-and-biases and in the reasoning research), S1 may generate quick automatic *non-normative* responses, while S2 may or may not intervene in its role as monitor and critic to correct or override S1's response. The relation of this framework to the concepts of intuition, cognition and meta-cognition as used in the mathematics education research literature (e.g., Fischbein 1987; Stavy and Tirosh 2000; Vinner 1997) is elaborated in Leron and Hazzan (2006), and will not be repeated here.

Many of the non-normative answers people give in psychological experiments—and in mathematics education tasks, for that matter—can be explained by the quick and automatic responses of S1, and the frequent failure of S2 to intervene in its role as critic and monitor of S1. The bat-and-ball task is a typical example for the tendency of the insuppressible and fast-reacting S1 to "hijack" the subject's attention and lead to a non-normative answer (Kahneman 2002). Specifically, the salient features of the problem cause S1 to "jump" automatically and immediately with the answer of 10 cents, since the numbers one dollar and 10 cents are salient, and since the orders of magnitude are roughly appropriate. For many people, the effortful and slow moving S2 is not alerted, and they accept S1's output uncritically, thus in a sense "behave irrationally" (Stanovich 2004). For others, S1 also immediately had jumped with this answer, but in the next stage, their S2 interfered critically and made the necessary adjustments to give the correct answer (5 cents).

Evans (2009) offers a slightly different view—called *default-interventionist* approach—of the relations between the two systems. According to this approach, only S1 has access to all the incoming data, and its role is to filter it and submit its "suggestions" for S2's scrutiny, analysis and final decision. In view of the huge amount of incoming information the brain constantly needs to process, this is a very efficient way to act because it saves the "expensive" working memory resources that S2 depends on. On the other hand, it is error-prone because the features that S1 selects are the most accessible but not always the most essential. In the bat-and-ball phenomenon, according to this model, the features that S1 has selected and submitted to S2 were the salient numbers 10 cents and 1 dollar and their sum, but the condition about the difference has remained below consciousness level. Even though S2 has the authority to override S1's decision, it may not do it due to lack of access to all the pertinent data.

Evolutionary psychologists, who study the ancient evolutionary origins of universal human nature, stress that the way S1 worked here, namely coming up with a very quick decision based on salient features of the problem and of rough sense of what's appropriate in the given situation, would be adaptive behavior under the natural conditions of our ancestors, such as searching for food or avoiding predators (Buss 2005; Cosmides and Tooby 1997; Tooby and Cosmides 2005). Gigerenzer (2005; Gigerenzer et al. 1999) claims that this is a case of *ecological rationality* being fooled by a tricky task, rather than a case of irrationality. These researchers and others also question whether the norms from logics, mathematics and probability are the correct norms for judging people's "rationality". Recent research, however, has shed new light on the question of what norms are appropriate for judging people's rationality. De Neys and Franssens (2009), using subtle indirect experimental techniques from memory research, have shown that people who give non-normative answers experience a conflict and are trying (albeit unsuccessfully) to inhibit their biased beliefs and stereotypes. Here is a brief summary of their conclusions:

> [These] results help to sketch a less bleak picture of human rationality. [...]
> [The present findings] establish that people are far more logical than their answers suggest. Although people's judgments are often biased, they are no mere intuitive, illogical thinkers who disregard normative considerations. All reasoners try to discard beliefs that conflict with normative considerations. The problem is simply that not everyone manages to complete the process. The inhibition findings have important implications for the status of logic and probability theory as normative standards. [...]
> The fact that people tried to block the intuitive beliefs when they conflicted with the traditional norms not only implies that people know the norms but also that they judge them to be relevant. If people did not believe that base-rates or logical validity mattered, they would not waste time trying to block the conflicting response. People might not always manage to adhere to the norm but they are at least trying to and are clearly not simply discarding it or treating it as irrelevant. This should at least give pause for thought before rejecting the validity of the traditional norms. (De Neys and Franssens 2009, p. 56)

In other words, people somehow sense when their beliefs (or intuitions, or stereotypes) are in conflict with the norms of logic or probability, and are trying to inhibit these biases. These new findings lend more theoretical support to the main idea of this article: If people sense the clash between their intuitive response and the normative one, they would be more motivated to try to reconcile the two, which is precisely what our bridging idea is meant to accomplish.

The seemingly paradoxical phenomenon (mentioned above) that more difficult task formulations actually enhance performance is also well explained by DPT. Since the subjects' S2 does possess the necessary mathematical knowledge, all that is required to solve the problem correctly is suppressing S1 and activating S2, which is exactly the effect of these added complications.

It is important to note that skills can migrate between the two systems. When a person becomes an expert in some skill, perhaps after a prolonged training, this skill may become S1 for this person. For example, speaking French is an effortful S2 behavior for many native English speakers, requiring deep concentration and full engagement of working memory resources. For experienced speakers who have lived in France for several years, however, speaking French may become an S1 skill which they can perform automatically while their working memory could be engaged in deep discussion on, for example, research in mathematics education. With enough training, similar migration can be observed in other skills such as reading, the multiplication table, or solving quadratic equations.

2.2 The Medical Diagnosis Problem (MDP)

In Sect. 3, we will demonstrate the theoretical, design-oriented, and experimental implications of the cognitive and educational challenges discussed in the introduction. We will do this mainly by focusing on one famous task concerning statistical thinking, which has been studied extensively by cognitive psychologists: the Medical Diagnosis Problem (MDP) and the related "base-rate neglect" fallacy. Here is a standard formulation of the task and data, taken from Samuels et al. (1999).

> Casscells et al. (1978) presented the following problem to a group of faculty, staff and fourth-year students at Harvard Medical School.
>
> > If a test to detect a disease whose prevalence is 1/1000 has a false positive rate of 5 %, what is the chance that a person found to have a positive result actually has the disease, assuming that you know nothing about the person's symptoms or signs? ___%
>
> Under the most plausible interpretation of the problem, the correct Bayesian answer is 2 %. But only 18 % of the Harvard audience gave an answer close to 2 %; 45 % of this distinguished group completely ignored the base-rate information and said that the answer was 95 %.

This task is intended to test what is usually called *Bayesian thinking*: how people update their initial statistical estimates (the *base rate*) in the face of new evidence. In this case, the initial estimate (the prevalence of the disease in the general population) is 1/1000, the new evidence is that the patient has tested positive (together with the 5 % false-positive rate[3]), and the task is intended to reveal how subjects will update their estimate of the probability that the patient actually has the disease. *Base-rate neglect* reflects the widespread fallacy of ignoring the base rate, instead of simply subtracting the false-positive rate (5 %) from 100 %. Indeed, it is not at all intuitively clear why the base rate should matter, and how it could be taken into the calculation.

In college-level courses on probability, a more comprehensive version of the MDP is often considered, where a nonzero *false-negative* rate is added. Since the statistical misconceptions (especially base-rate neglect) are already revealed in the simpler version considered here, most cognitive researchers did not find it necessary to engage subjects with the harder version.

A formal solution to the various versions of the task is based on Bayes' theorem (DeGroot 1985, p. 66), but there are many complications and controversies involving mathematics, psychology and philosophy, concerning the interpretation of that theorem. Indeed, this debate—"Are humans good intuitive statisticians after all?" (Cosmides and Tooby 1996) is a central issue in the great rationality debate. Or put in the words of Barbey and Sloman (2007), who present a comprehensive discussion of the MDP and the surrounding controversy:

> [The] extent to which intuitive probability judgment conforms to the normative prescriptions of Bayes' theorem has implications for the nature of human judgment. (p. 242)

[3]This means that 5 % of the healthy people taking the test will test positive.

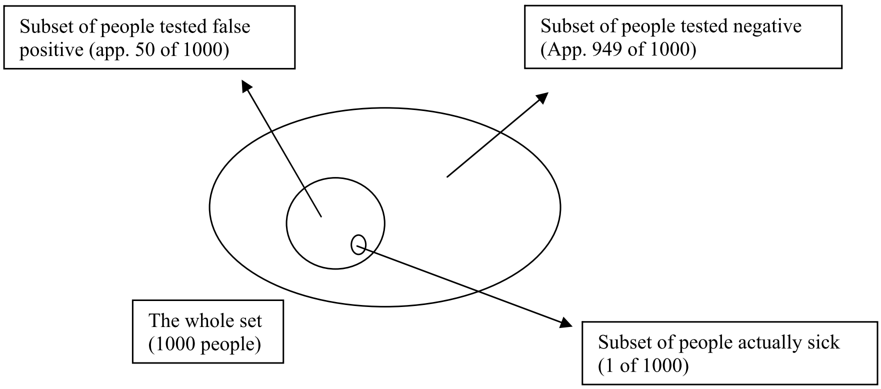


Fig. 1 Nested subsets in the medical diagnosis problem (for 1000 people)

Here is a simple intuitive solution for the MDP, bypassing Bayes' theorem:[4] Assume that the population consists of 1000 people and that all have taken the test (Fig. 1). We know that one person will have the disease (because of the base rate) and will test positive (because no false negative rate is indicated). In addition, 5 % of the remaining 999 healthy people (approximately 50) will test false-positive—a total of 51 positive results. Thus, the probability that a person who tests positive actually has the disease is $1/51$, which is about 2 %.

Reflecting on the foregoing argument, we note that its remarkable simplicity has been achieved via two quite general methods. First, by imagining a village with a population of 1000—a *generic example* representing *any* village—we have utilized the simplifying tool of *generic proof* (e.g., Movshovitz-Hadar 1988; Leron and Zaslavsky in press). Secondly, the probability calculations have been much simplified by using a diagram of the various subsets and their nesting relationships (Fig. 1), and by calculating with *natural frequencies* (numbers of persons) instead of probabilities (fractions between 0 and 1). Not coincidentally, the two simplifying concepts that came up in this reflection—the nested subsets relationships and natural frequencies—have figured prominently in the debate among cognitive psychologists concerning the MDP, and are further discussed below.[5]

To appreciate the gains and losses of the foregoing approach, we now remove these simplifications and translate the previous argument to a more general and rigorous (though less elementary and intuitive) solution. It turns out that our simplified argument was implicitly mimicking a calculation by the standard *conditional probability* formula $\Pr(A|B) = \Pr(AB)/\Pr(B)$ (DeGroot 1985, p. 57), which in our case can be written as follows:

[4] In our experience, this solution can be easily understood even by junior high school students.

[5] A third kind of simplification—rounding off 999 to 1000 and $1/51$ to $1/50$—is more specific to the numbers given here, and is not similarly generalizable.

$$p = \Pr(\text{has disease}|\text{tested positive})$$
$$= \Pr(\text{has disease } \textit{and} \text{ tested positive}) / \Pr(\text{tested positive}).$$

Thus, our previous argument "we know that one person will have the disease (because of the base rate) and will test positive (because no false negative rate is indicated)", is now used to show that the *numerator* in the conditional probability formula is equal to 1/1000. The *denominator*, namely the fraction of people who tested positive, is seen by the above argument (no rounding off this time) to be (1 + 5 % of 999)/1000, or 50.95/1000. Thus we have $p = (1/1000)/(50.95/1000)$ and, multiplying numerator and denominator by 1000, we get: $p = 1/50.95 = 1.96\ \%$.

Note that this calculation is not limited to a population of 1000. The 1000 appearing in the calculation is there due to the given 1/1000 base rate.

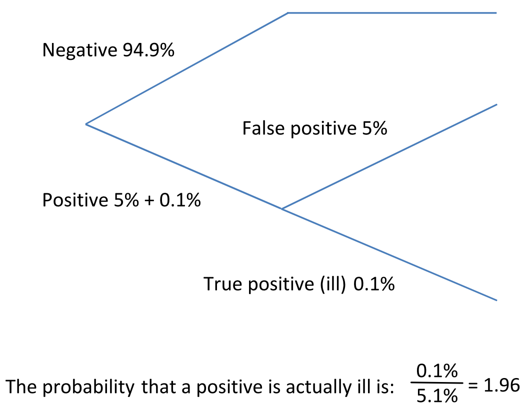
To complete this discussion of possible routes to solving the MDP, we mention that there is also an approach to conditional probability using a "tree diagram", which we will not pursue here, since none of our subjects used it. Figure 2 represents such a tree diagram for our case.

Researchers with evolutionary and ecological orientation (Cosmides and Tooby 1996; Gigerenzer et al. 1999) claim that people are "good statisticians after all" if only the input and output is given in "natural frequencies" (integers instead of fractions or percentages):

> [...] we will explore what we will call the "frequentist hypothesis"—the hypothesis that some of our inductive reasoning mechanisms do embody aspects of a calculus of probability, but they are designed to take frequency information as input and produce frequencies as output. (Cosmides and Tooby 1996, p. 3)

Evolutionary psychologists theorize that the brains of our hunter–gatherer ancestors developed such a module because it was vital for survival and reproduction, and because this is the statistical format that people would naturally encounter under those conditions. The statistical formats of today, in contrast, are the result of the huge amount of information that is collected, processed and shared by modern societies with modern technologies and mass media.

Fig. 2 Tree diagram for the MDP



Indeed, Cosmides and Tooby (1996) have replicated the Casscells et al. (1978) experiment, but with natural frequencies replacing the original fractional formats and the base-rate neglect has all but disappeared:

> Although the original, non-frequentist version of Casscells et al.'s medical diagnosis problem elicited the correct Bayesian answer of 2 % from only 12 % of subjects tested, pure frequentist versions of the same problem elicited very high levels of Bayesian performance: an average of 76 % correct for purely verbal frequentist problems and 92 % correct for a problem that requires subjects to construct a concrete, visual frequentist representation. (Cosmides and Tooby 1996, p. 58)

These results, and the evolutionary claims accompanying them, have been consequently challenged by other researchers (Evans 2006; Barbey and Sloman 2007). In particular, Evans (2006, 2009) claims that what makes the subjects in these experiments achieve such a high success rate is not the frequency format per se, but rather a problem structure that cues explicit mental models of nested-set relationships (as in Fig. 1 above). However, the fresh perspective offered by evolutionary psychology has been seminal in re-invigorating the discussion of statistical thinking in particular, and of cognitive biases in general. The very idea of the frequentist hypothesis, and the exciting and fertile experiments that it has engendered by supporters and opponents alike, would not have been possible without the novel evolutionary framework. Here is how Samuels et al. (1999, p. 101) summarize the debate:

> But despite the polemical fireworks, there is actually a fair amount of agreement between the evolutionary psychologists and their critics. Both sides agree that people do have mental mechanisms which can do a good job at Bayesian reasoning, and that presenting problems in a way that makes frequency information salient can play an important role in activating these mechanisms.

Barbey and Sloman (2007) survey the vast literature on the MDP, categorize the theoretical stances adopted by the various researchers, and test how well they stand the experimental evidence. Here is a brief summary:

> Explanations of facilitation in Bayesian inference can be grouped into five types that can be arrayed along a continuum of cognitive control, from accounts that ascribe facilitation to processes that have little to do with strategic cognitive processing to those that appeal to general purpose reasoning procedures. (p. 242)

Their conclusion after a long and detailed analysis of the data and the relevant literature is in line with Evans' quote above:

> The evidence instead supports the nested sets hypothesis that judgmental errors and biases are attenuated when Bayesian inference problems are represented in a way that reveals underlying set structure, thereby demonstrating that the cognitive capacity to perform elementary set operations constitutes a powerful means of reducing associative influences and facilitating probability estimates that conform to Bayes' theorem. An appropriate representation can induce people to substitute reasoning by rules with reasoning by association. (p. 253)

2.3 Connections to Previous Research in Mathematics Education

There are two strands of research in mathematics education which are particularly relevant to our discussion. First, there is the research on intuitive thinking, notably Fischbein (1987) and Stavy and Tirosh (2000). A survey of this research and its similarity and difference with DPT has been presented in Leron and Hazzan (2006), and need not be repeated here. The second relevant strand is research on statistics (or probability) education, which will be surveyed here only briefly, since probability enters our topic only incidentally. As we explain in the introduction, this is neither a paper about statistical thinking nor about the MDP. The MDP serves us in the role of a case study, to investigate the clash between intuition and analytical thinking. It has been selected because of the extensive research on it by psychologists, and, in principle, we could have just as well chosen another problem from a different topic area. For a more extensive review of the literature on statistical education, see, for example, Abrahamson (2009).

As mentioned in the introduction, we hold the position that one cannot just eliminate or replace intuitions, and that it is, in fact, possible and desirable to build new knowledge from existing one, even when the latter deviates from the desired analytical norms. Smith et al. (1993) argue convincingly against the idea that misconceptions can—or should—be "replaced" or "confronted" as an instructional strategy:

> Replacement—the simple addition of new expert knowledge and the deletion of faulty misconceptions—oversimplifies the changes involved in learning complex subject matter. [. . .] Literal replacement itself cannot be a central cognitive mechanism [. . .], nor does it even seem helpful as a guiding metaphor [. . .] Evidence that knowledge is reused in new contexts—that knowledge is often refined into more productive forms—and that misconceptions thought to be extinguished often reappear [. . .] all suggest that learning processes are much more complex than replacement suggests.
> [. . .]
> Instruction designed to confront students' misconceptions head-on [. . .] is not the most promising pedagogy. It denies the validity of students' conceptions in all contexts; it presumes that replacement is an adequate model of learning; and it seems destined to undercut students' confidence in their own sense-making abilities. Rather than engaging students in a process of examining and refining their conceptions, confrontation will be more likely to drive them underground. [. . .] The goal of instruction should be not to exchange misconceptions for expert concepts but to provide the experiential basis for complex and gradual processes of conceptual change. (pp. 153–154)

Abrahamson (2009) is a closely related article, and is in several ways complementary to ours. Both articles are motivated by respect for learners' pre-conceptions and intuitions, and both use bridging to help learners grow the required normative knowledge out of their existing knowledge. The main difference is that Abrahamson discusses bridging as part of an integrative teaching-learning system (as it should clearly be in real classroom application), while we are studying its purely cognitive aspects. A second difference is that Abrahamson goes more deeply into the specifics of statistical thinking, while we stay at the level of general cognitive processes, using the statistical task only as an example. Thus Abrahamson's analysis is closer to classroom reality, while ours focuses more sharply on the cognitive aspects of bridging tasks.

3 The Educational Challenge as Design Issue

3.1 Confronting the Educational Challenge with Bridging Tasks

To demonstrate what we mean by the title "the educational challenge as design issue", we again start with a puzzle, this time the famous string-around-the-earth puzzle (dating back to 1702).

> Imagine you have a string tightly encircling the equator of a basketball. How much extra string would you need for it to be moved one meter from the surface at all points? Hold that thought, and now think about a string tightly encircling the Earth—making it around 40,000 kilometers long. Same question: how much extra string would you need for it to be one meter from the surface at all points?

Everybody seems to feel strongly that the Earth would need *a lot* more extra string than the basketball. The surprising answer is that they both need the same amount: 2π, or approximately 6.28 meters. (If R is the radius of any of them, then the extra string is calculated by the formula $2\pi(R+1) - 2\pi R = 2\pi$.)

As with the bat-and-ball puzzle, *this surprise is what we are after*, for it tells us something important about how the mind works, which is why cognitive psychologists are so interested in such puzzles. This time, however, in addition to the cognitive challenge, we are also concerned with the corresponding *educational challenge*. Suppose as before that you present this puzzle to your math class. Being a seasoned math teacher, you first let them be surprised; that is, you first let them do some guessing, bringing out the strong intuitive feeling that the required additional string is small for the basketball but huge for the earth. Then you have them carry out the easy calculation (as above) showing that—contrary to their intuition—the additional string is actually quite small, and is in fact independent of the size of the ball. Now, given the classroom situation just described, here again is the educational challenge: *As teachers and math educators, what do we do next?* As we have said in the introduction, we want to avoid the "default" conclusion that students should not trust their intuition, and should abandon it in the face of conflicting analytical solution. We also want to help students overcome the uncomfortable situation, whereby their mind harbors two conflicting solutions, one intuitive but wrong, the other correct but counter-intuitive. In other words, we believe with Smith et al. (1993) that the new analytical knowledge should not *replace* the existing intuitive knowledge but should rather *grow out* of it.

Papert's (1993/1980, pp. 146–150) answer to this educational challenge is simple but ingenious: Just imagine a cubic earth instead of a spherical one! Now follow in your mind's eye the huge square equator with two strings, one snug around its perimeter and the other running in parallel 1 meter away (Fig. 3). Then you can actually *see* that the two strings have the same length along the sides of the square (the size doesn't matter!), and that the only additional length is at the corners. In addition, you can now *see* why the extra string should be 2π: It is equal to the perimeter of the small circle (of radius 1 meter) that we get by joining together the 4 quarter-circle sectors at the corners.

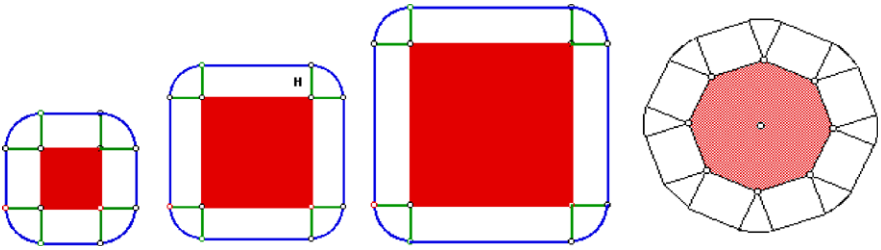


Fig. 3 A string around a square Earth—and around an octagonal one

The final step in Papert's ingenious construction is to bridge the gap between the square and the circle with a chain of perfect polygons, doubling the number of sides at each step. The next polygon in the chain after the square would be an octagon (Fig. 3, right). Here we have 8 circular sections at the corners, each half the size of those in the square case, so that they again can be joined to form a circle of exactly the same size as before. This demonstrates that doubling the number of sides (and getting closer to a circle) leaves us with the same length of extra string.

Can this beautiful "bridging" example be generalized? When intuitive and analytical thinking clash, can we always design such "bridging tasks" that will help draw them closer? How should theory, design and experiment collaborate in the pursuit of this goal?

We will soon turn to the medical diagnosis problem MDP to examine these questions in more detail, but let us first look at one more elementary example. This example shows that the idea of bridging can be effective not only in puzzle-like cognitive challenges, but also in mainstream curriculum topics and in concept development, where it is often the case that a certain topic is hard to learn not because it is very complicated or abstract, but because the idea to be learned clashes with learners' existing intuitions (possibly due to previous learning).

Consider the well-known misconception among young learners, who often think that the decimal 3.14 is bigger than the decimal 3.5 "because 14 is bigger than 5" (an intuition carried over from the natural numbers; see Resnick et al. 1989). Here it is easy to think of a bridging task that would help reconcile the two conflicting perspectives. For example, without any further teaching or explaining, we can pose the following variant: "*Which number is bigger*: 3.14 or 3.50?" Then we can go back to the original task.

This idea is not original, and is, in fact, a famous "trick" among teachers, but it is most often taught as a rule: "add zeroes at the end until the two decimals have the same number of digits after the decimal point". But note that there is an important difference between the use of this idea as a rule (signaling to the student "memorize and follow") and its use as a bridging task (signaling "use the analogy to engage your common sense"). In a way, we are not only making the new idea more intuitive, but are also helping the learners stretch their intuition towards the desired mathematical idea (Wilensky 1991; Smith et al. 1993; Abrahamson 2009).[6]

While cognitive psychologists are interested in this kind of tasks and in people's behavior when confronted with them, the goals of researchers in cognitive psychology and in mathematics education are quite different. The former are mainly interested in documenting and understanding people's behavior "in the wild", and what it reveals about the workings of the human mind. Thus exposing people's fallible intuitions is for them a goal by itself. Mathematics education researchers, in contrast, want to know what can be done about it. How can we "stretch" our students' intuitive thinking into building more sophisticated knowledge structures? This approach is related to the recent trend on viewing mathematics education as *design science*. There is a hint in our work, though only in a very preliminary way, of a typical theory-design-test cycle of design research:

> [...] we define design research as a family of methodological approaches in which instructional design and research are interdependent. On the one hand, the design of learning environments serves as the content for research, and, on the other hand, ongoing and retrospective analyses are conducted in order to inform the improvement of the design. This type of research involves attempting to support the development of particular forms of learning and studying the learning that occurs in these designed settings. (Cobb and Gravemeijer 2008, p. 68)

Indeed, we start with a hypothesis based on a theoretical framework, use it to design an educational artefact according to certain desiderata in the form of pre-assigned constraints, then test it through an experiment to see if our product indeed satisfies the pre-assigned constraints. In our case, the artefact is the new bridging task, and the constraints are its relation to the original task as detailed below.

3.2 A Case Study: Bridging the Medical Diagnosis Problem (MDP)

The extensive data on base rate neglect in the MDP (leading to the 95 % answer) demonstrates the counter-intuitive nature of the analytical solution, as in the case of the string around the earth. As math educators, we are interested in helping students build bridges between the intuitive and analytical perspectives, hopefully establishing peaceful co-existence between these two modes of thinking. As in Papert's example, achieving such reconciliation involves a design issue: Design a new *bridging task*, which is logically equivalent to, but psychologically easier than the given task. (Compare Clements' (1993) "bridging analogies" and "anchoring intuitions" in physics education.)

From the extensive experimental and theoretical research on the MDP in psychology, we were especially influenced in our design efforts by the *nested subsets hypothesis* (see Fig. 1):

---

[6]These two modes of helping the students are termed, respectively, *bridging down* and *bridging up* in a forthcoming paper, where they are treated more thoroughly.

> All this research suggests that what makes Bayesian inference easy are problems that provide direct cues to the *nested set* relationships involved [...]
> It appears that heuristic [S1] processes cannot lead to correct integration of diagnostic and base rate information, and so Bayesian problems can only be solved analytically [i.e., by S2]. This being the case, problem formats that cue construction of a single mental model that integrates the information in the form of nested sets appears to be critical. (Evans 2006, p. 391)

Indeed, it is not easy to form a mental representation of the subsets of sick and healthy people, and even less so for the results of the medical test. Mental images of people all look basically the same, whether they are sick or healthy, or tested positive or negative. The task of finding a more intuitive version of the MDP has thus been operationalized to finding a task which will "cue construction of a *single mental model* that integrates the information in the form of nested sets" (ibid).

Based on this theoretical background, we formulated three *design criteria* for the new task (which would also serve as testable predictions).

1. *Intuitively accessible*: The bridging task we will design will be easier ("more intuitive") than the original MDP (i.e., significantly more people—the term is used here in a qualitative sense—will succeed in solving it correctly).
2. *Bridging function*: Significantly more people will solve the MDP correctly, with no additional instruction, after having solved the new task.
3. *Nested subsets hypothesis*: Base rate neglect will be significantly reduced.

Note that the first two design criteria pull the new task in opposite directions. Criterion 1 (turning the hard task into an easy one) requires a task which is sufficiently different from the original one, while criterion 2 (the bridging function) requires a task that is sufficiently similar to the original one. The new task, then, should be an *equilibrium point* in the "design space"—sufficiently different from the original task but not *too* different.

Armed with these criteria, we set out on the search for the new bridging task. As a preliminary step we changed slightly the original formulation of the MDP, aiming to make it easier to understand without making it easier to solve. In our judgment, the original version was unduly cryptic, and we were interested in people's statistical thinking, not their reading comprehension. Here is our version of the MDP.

MDP In the country of Oz, 1 person in 1000 has the disease Atshi. A test to detect the disease has a false positive rate of 5 %, which means that 5 % of the healthy people are falsely tested positive. What is the chance that a person who tested positive actually has the disease? ___%

We then embarked on a long process of trial and error, intermediate versions, partial successes and failures, together with pilot experiments, until we have finally come up with the *Robot-and-Marbles Problem* (RMP), which we felt had a good chance of satisfying the design criteria and withstanding the empirical test. The RMP is based on the idea of replacing sick and healthy people in the population by red and green marbles in a box. The medical test is then replaced by a color-detecting robot, which can distinguish between red and green marbles via a color sensor. The sensor, however, is not perfect and 5 % of the green marbles are falsely identified as red (corresponding to the 5 % healthy people in the MDP who are falsely diagnosed as sick). We also decided to make the action of the robot on the marbles more vividly imaginable by actually describing the process, not just the result. A final step in the design of the new problem was to slightly change the numbers from the original MDP, in order to make the connection less obvious. According to the bridging criterion, our subjects who solved the RMP first, should then solve more successfully the MDP. For this to happen, they would first need to recognize the *similarity* between the two problems, and we didn't want to make this too obvious by using the same numbers.

Here then is the final product of our design efforts, the version that would actually be put to the empirical test, to see whether the design criteria have been satisfied.

The Robot-and-Marbles Problem (RMP) In a box of red and green marbles, 2/1000 of the marbles are red. A robot equipped with green-marble detector with a 10 % error rate (10 % green marbles are identified as red), throws out all the marbles which it identifies as green, and then you are to pick a marble at random from the box. What is the probability that the marble you have picked would be red?

Note The correct answer to the RMP, even with the new numbers, is still approximately 2 %, since the prevalence and the false-positive rate appear in the calculation in the numerator and the denominator, respectively, so that doubling both leaves the answer unchanged.

3.3 The Experiment

The participants in the experiment were 128 students in a methodology course, studying towards an M.A. degree in Educational Psychology at a Danish university. The participants had no special background in mathematics or statistics, and had not been prepared for the mathematical nature of the task. They were all informed—and have accepted—that they took part in an experiment, which could thus fit well into the subject matter (methodology) of the course. All the participants were assigned the two tasks—the medical diagnosis problem (MDP) and the robot-and-marbles problem (RMP)—and were given 5 minutes to complete each task. (In a pilot experiment, we found that 5 minutes were enough both for those who could solve the problem and those who couldn't.) The subjects were assigned randomly into two groups of 64 students each. The order of the tasks was MDP first and RMP second for one group (called here the *MR group*), and the reverse order for the second group (the *RM group*). The results of the RM group were clearly our main interest, the MR group serving mainly as control.

The results are collected in Table 1, from which it can be seen that the design criteria have been validated and the predictions confirmed. The results of the Harvard

Table 1 Numbers of responses in the various categories

	RM group (robot first)		MR group (medical diagnosis first)		
	RMP 1st	MDP 2nd	MDP 1st	RMP 2nd	Harvard
Correct	31 (48 %)	17 (25 %)	8 (12 %)	20 (31 %)	18 %
Base-rate neglect	1 (2 %)	12 (19 %)	22 (34 %)	4 (6 %)	45 %
Incorrect other	32	35	34	40	
Total	64	64	64	64	

experiment (described at the beginning of Sect. 2.2) are added at the right-most column for comparison. Summarizing the main points, the following conclusions for the RM group (with comparative notes in parentheses) can be read off the table.

1. The RMP succeeded in its role as a bridge between intuitive and analytical thinking: 48 % of the subjects in the RM group solved it correctly, compared to 18 % success on the MDP in the original Harvard experiment and 12 % in our MR group.
2. The RMP succeeded in its role as a stepping stone for the MDP: More than 25 % solved the MDP, without any instruction, when it followed the RMP, again compared to 18 % in the original Harvard experiment and 12 % in our MR group.
3. The notorious base-rate neglect has all but disappeared in the RMP: It was exhibited by only 1 student out of 64 in the RM group and 4 out of 64 in the MR group, Compared to 45 % on the MDP in the original Harvard group and 34 % in our MR group.
4. Remarkably, the MDP, when given first, does not at all help in solving the RMP that follows. Worse, the MDP gets in the way: The table shows 48 % success on the RMP alone, vs. 31 % success on the RMP when given after the MDP.
5. Even though the performance on the RMP and the MDP has greatly improved in the RM group, still the largest number of participants appear in the “incorrect other” category. This category consists of diverse errors which do not directly relate to the MDP, including (somewhat surprisingly for this population) many errors concerning misuse of percentages.

3.4 Discussion

The classification of the answers was not easy or simple, and we had to make many ad-hoc decisions on the way. Unlike the researchers in cognitive psychology, who typically look only at the answers, we have also examined the students’ scribbling on the answer sheets, in an attempt to understand their way of reasoning, and have tried to take this information into account when classifying their answers.

An unexpected obstacle was the difficulties many participants had in dealing with percentages. For example, one student has arrived at the (roughly correct) answer

"2 out of 102", which she then converted to what would be her final (and incorrect) answer, 1.02 %. We have decided to classify this answer as correct, since this student has shown correct statistical reasoning (which after all had been the target of this experiment), while erring in the use of percentages (which we considered unfortunate but irrelevant in this context).

Even though the bridging task did satisfy the design criteria, the relationship between the two tasks are far from settled. Looking at the numbers in the first row of the table, three issues are outstanding, which may require further research to understand more fully.

First, a small number of the students managed to solve the MDP but not the RMP. The numbers here are too small to draw any definite conclusions, and this might be an incidental phenomenon. In informal discussions that we held with some participants after the experiment, we found out that some of the students who were able to solve the MDP but not the RMP were nurses, who claimed that they were helped by the familiar medical context.

Second, there were 19 students who did answer the RMP 1st correctly but still couldn't solve the MDP 2nd. Since the two tasks are really isomorphic (thus in a sense "the same"), it seems that these students simply did not notice the connection: The fact that the two tasks are isomorphic is only helpful if you *notice* this connection.

Third, the fraction of students who solved successfully either the RMP (31/64) or the MDP (17/64) is still quite small, even with the bridging task. Though this may seem disappointing, it is hardly surprising. The bridging task is meant to help students reconcile the intuitive and analytical solutions when they already know both (see the string-around-the-earth example), but it cannot miraculously make all of them solve a difficult task in probability *without any explicit instruction*. Because the emphasis in this article was on the cognitive role of the bridging task, we refrained from talking about instructional interventions, but obviously a good combination of the two tools would be needed to obtain better results (Abrahamson 2009).

The following comments indicate that the difficulty of the MDP (and even the RMP) may have affective sources as well as the much discussed cognitive ones. In fact, quite a few (about 10 % of the participants) devoted the 5 minutes allotted for the solution to express in writing their negative feelings about mathematics. There were no similar positive comments, presumably because those with positive feelings just went ahead and worked on the solution. We bring here a sample.

- It's mathematics, and my brain power shuts down when it sees a % sign.
- Wow, this is a really hard! My thoughts go back to my math lessons in elementary school. I feel a terrible frustration! I am good at so many other things as language, culture, people and psychology, but absolutely not mathematics.
- Was confused in the first place because I HATE mathematics, but chose to think and read.
- It's total darkness for me… I block… miss the motivation to figure out this task
- NOOO, hell! I don't know how many marbles I get!
- I have known for a long time that I was really mathematically challenged. Already before I really started to look at the task, my brain blocked totally for any solution.

The reason why I give up in advance can be that I am afraid of making mistakes. I am unable to solve such tasks because I find them totally uninteresting.

- I am firmly fixed in my mental mathematical block. I recall the sweating problem solving in mathematics and my usual displacement maneuvers. I recall the statement that if you stop drawing at 10 years old, then you will always remain at 10. Just now I am back in grade 9, with the impulse of running away.

The role of affective factors in problem solving is an important issue that deserve a more extensive treatment. This, however, is not the focus of the present article, and would take us too far away from our main focus.

One more possible extension of our method concerns biases that arise not from cognitive biases (as in the bat-and-ball puzzle or the MDP), but from social biases or *stereotypes*, the kind of biases that has been studied extensively by social psychologists. We look at one such task, due originally to Kahneman and Tversky, and discussed further in De Neys and Franssens (2009):

> Participants [...] were asked to solve problems that were modeled after Kahneman and Tversky's [...] base-rate neglect problems. Consider the following example:
> In a study, 100 people were tested. Jo is a randomly chosen participant of this study. Among the 100 participants there were 5 men and 95 women.
> Jo is 23 years old and is finishing a degree in engineering. On Friday nights, Jo likes to go out cruising with friends while listening to loud music and drinking beer.
> What is most likely:
>
> (a) Jo is a man.
> (b) Jo is a woman.
>
> Given the size of the two groups in the sample, it will be more likely that a randomly drawn individual will be a woman. Normative considerations based on the group size or base-rate information cue response (b). However, many people will be tempted to respond (a) on the basis of stereotypical beliefs cued by the description. [Once again,] normative considerations will conflict with our beliefs, and sound reasoning requires inhibition of the compelling but erroneous belief-based response.
> One can easily construct no-conflict or control versions of the base-rate problems. In the no-conflict version the description of the person will simply be composed of stereotypes of the larger group [...] Hence, contrary to the classic problems, base-rates and description will not conflict and the response can be rightly based on the beliefs cued by the description without any need for inhibition. (p. 50)

In our view, the suggestion in the last paragraph—reversing the description so that the stereotype would lead to the normative answer—misses the point of a good bridging task because it goes along with the stereotypical bias rather than challenging it. We conjecture that a good bridging problem will be one which is logically equivalent to the original one, but where the stereotypical content has been removed rather than reversed. Specifically, we conjecture (and intend to show in a future experiment) that the following problem could serve as an effective bridge: Assume that you have in a box 100 marbles, 5 of which are blue and 95 red. If you pick one marble from the box at random, what is most likely: (a) the marble is blue; (b) the marble is red?

By saying that this is an effective bridge, we mean that it should satisfy the three design criteria given above: First, it should be intuitively accessible, that is, significantly more people will solve it than the original task, and second (the bridging function), significantly more people will solve the original task, without any further instruction, after having solved our task.

It is important to note that we are not advocating eliminating problems that invoke stereotypes. On the contrary, we want to help students learn how to overcome stereotypes in solving mathematical problems (and perhaps, indirectly, in life in general). The new task does not “replace” the original task; it only serves as a bridge to get there more safely.

4 Conclusion

In this article, we have seen how the dual-process theory from cognitive psychology highlights and help explain the power of intuitive thinking and its uneasy relation with analytical thinking. We have used the extensively-researched Medical Diagnosis Problem as a case study for discussing the gap between intuition (System 1) and analytical thinking (System 2), and for developing design principles to bridge this gap. Our design is aimed not at making the task easier, but rather at creating a new bridging task, that will serve as stepping stone to help students climb from their current intuitive understanding towards the more difficult and often counter-intuitive analytical solution. It is our belief that bridging the gap between intuition and analytical thinking (in research, curriculum planning, learning environments, teaching methods, work with teachers and students) is a major step towards improving student conceptual understanding in mathematics.

Above all, we believe that intuitive thinking, or common sense, is a powerful tool in people’s everyday behavior, thinking and reasoning, and we are concerned that for many students common sense is not part of their mathematical tool box. By helping students build bridges between their intuitive thinking and the required analytical thinking, we hope that common sense will regain its important role in their mathematical thinking. We hope to replace the tacit message “common sense (or intuition) is risky and should be ignored or overcome” with the new message, “common sense is a powerful tool that we are all naturally equipped with, and we should use it extensively in learning mathematics, while at the same time we should exercise suitable care in monitoring, modifying and extending it when necessary.”

The current study is but a first step in what could become a major research and development project, and we mention two possible directions for continuing and extending the current study. One, analyze more case studies and try to describe more fully the characteristics of a good bridging task. From the current examples we have examined (three in this paper and a few more unpublished ones), it seems that good bridging tasks can be constructed from the given tasks by concretizing and discretizing the objects involved and their relations (e.g., square earth instead of round one, green and red pebbles instead of healthy and sick people). The explanation would be in accord with Evans’ claim, since these new objects would seem to enhance the subject’s ability to faithfully represent the problem’s constraints in a single mental model. In other cases (such as the decimal comparison task, where the intuition comes from their knowledge of integers), the bridging task would be closer to students' previous knowledge, rather than to their universal intuitions.

The second extension would be to develop and study bridging tasks not only in puzzle-like cognitive challenges, but also in mainstream curriculum topics and in concept development, as has been demonstrated above for the decimal comparison task. This would be helpful in situations where misconceptions arise in learning a topic not because it is particularly complex or abstract, but rather because it goes against existing intuitions (that may be either part of "human nature" or acquired in previous learning).

Based on the above examples and analysis, and indulging in a bit of over-optimism, here are some of the developments in the educational system we'd like to see happen in the future:

– Map out the high school curriculum (and beyond) for components that could build on natural thinking and parts that would clash with it. For example, which aspects of functions (or fractions, or proofs) are consonant or dissonant with natural thinking?
– Design curricula, learning environments, teaching methods that build from the power of natural thinking.
– Build a stock of puzzles and problems which challenge the intuition, and develop ways to work profitably with teachers and students on these challenges.

We wish to conclude with an even bigger educational challenge. If you ask mathematicians for examples of beautiful theorems, you will discover that many of them are *counter-intuitive*; indeed, that they are beautiful *because* they are counter-intuitive, *because* they challenge our natural thinking. Like a good joke, the beauty is in the surprise, the unexpected, the unbelievable. Like a world-class performance in classical ballet, sports, or a soprano coloratura, the beauty is (partly) in overcoming the limitations of human nature. Examples abound in the history of mathematics: The infinity of primes, the irrationality of $\sqrt{2}$, the equi-numerosity of even and natural numbers, the impossibility theorems (trisecting angles by ruler and compass, solving 5th-degree equations by radicals, enumerating the real numbers, Gödel's theorems). Recall, too, the joy of discovering that—contrary to your intuition—the extra length in the string-around-the-earth puzzle is quite small, and the beauty of Papert's cubic earth thought experiment.

Here, then, is the challenge: By all means, let us build on the power of natural thinking, but let us also look for ways to help our students feel the joy and see the beauty of going beyond it, or even against it. We can thus summarize the themes of this article with the following slogan: *The power of natural thinking, the challenge of stretching it, the beauty of overcoming it*.

References

Abrahamson, D. (2009). Orchestrating semiotic leaps from tacit to cultural quantitative reasoning—the case of anticipating experimental outcomes of a quasi-binomial random generator. *Cognition and Instruction*, *27*(3), 175–224.

Barbey, A. K., & Sloman, S. A. (2007). Base-rate respect: from ecological rationality to dual processes. *Behavioral and Brain Sciences*, *30*(3), 241–254.

Buss, D. M. (Ed.) (2005). *The handbook of evolutionary psychology*. New York: Wiley.

Casscells, W., Schoenberger, A., & Grayboys, T. (1978). Interpretation by physicians of clinical laboratory results. *The New England Journal of Medicine*, *299*, 999–1000.

Clements, J. (1993). Using bridging analogies and anchoring intuitions to deal with students' preconceptions in physics. *Journal of Research in Science Teaching*, *30*(10), 1241–1257.

Cobb, P., & Gravemeijer, K. (2008). Experimenting to support and understand learning processes. In A. E. Kelly, R. A. Lesh, & J. Y. Baek (Eds.), *Handbook of Design research methods in education* (pp. 68–95). London: Routledge.

Cosmides, L., & Tooby, J. (1997). *Evolutionary psychology: a primer*. Center for Evolutionary Psychology, University of California, Santa Barbara. Accessed on August 4, 2010 at http://www.psych.ucsb.edu/research/cep/primer.html.

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

DeGroot, M. H. (1985). *Probability and statistics* (2nd ed.). Reading: Addison-Wesley.

De Neys, W., & Franssens, S. (2009). Belief inhibition during thinking: not always winning but at least taking part. *Cognition*, *113*, 45–61.

Evans, St.B.T.J. (2006). The heuristic-analytic theory of reasoning: extension and evaluation. *Psychonomic Bulletin & Review*, *13*(3), 378–395.

Evans, St.B.T.J. (2009). How many dual-process theories do we need? One, two, or many? In *Evans and Frankish* (pp. 33–54).

Evans, St.B.T.J., & Frankish, K. (Eds.) (2009). *Two minds: dual processes and beyond*. London: Oxford University Press.

Fischbein, E. (1987). *Intuition in science and mathematics: an educational approach*. Dordrecht: Reidel.

Gigerenzer, G., Todd, P. M., & the ABC Research Group (1999). *Simple heuristics that make us smart*. London: Oxford University Press.

Gigerenzer, G. (2005). I think, therefore I err. *Social Research*, *72*(1), 1–24.

Kahneman, D., & Nobel Prize Lecture, D. (2002). Maps of bounded rationality: a perspective on intuitive judgment and choice. In T. Frangsmyr (Ed.), *Les Prix Nobel* (pp. 416–499). Also accessible at http://www.nobel.se/economics/laureates/2002/kahnemann-lecture.pdf.

Kahneman, D., & Frederick, S. (2005). A model of Heuristic judgment. In K. J. Holyoak & R. G. Morisson (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 267–294). Cambridge: Cambridge University Press.

Leron, U., & Hazzan, O. (2006). The rationality debate: application of cognitive psychology to mathematics education. *Educational Studies in Mathematics*, *62*(2), 105–126.

Leron, U., & Hazzan, O. (2009). Intuitive vs. analytical thinking: four perspectives. *Educational Studies in Mathematics*, *71*, 263–278.

Leron, U., & Zaslavsky, O. (in press). Generic proving: reflections on scope and method. *For the Learning of Mathematics*.

Movshovitz-Hadar, N. (1988). Stimulating presentations of theorems followed by responsive proofs. *For the Learning of Mathematics*, *8*(2), 12–19.

Papert, S. (1993/1980). *Mindstorms: children, computers and powerful ideas* (2nd ed.). New York: Basic Books.

Resnick, L., Nesher, P., Leonard, E., Magone, M., Omanson, S., & Peled, I. (1989). Conceptual bases of arithmetic errors: the case of decimal fractions. *Journal for Research in Mathematics Education*, *20*(1), 8–27.

Samuels, R., Stitch, S., & Tremoulet, P. (1999). Rethinking rationality: from bleak implications to Darwinian modules. In E. LePore & Z. Pylyshyn (Eds.), *What is cognitive science?* (pp. 74–120). Oxford: Blackwell Sci.

Song, H., & Schwarz, N. (2008). Fluency and the detection of misleading questions: low processing fluency attenuates the Moses illusion. *Social Cognition*, *26*(6), 791–799.

Smith, J. P., diSessa, A. A., & Roschelle, J. (1993). Misconceptions reconceived: a constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences*, *3*(2), 115–163.
Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: implications for the rationality debate. *Behavioral and Brain Sciences*, *23*, 645–726.
Stanovich, K. E. (2004). *The robot rebellion: finding meaning in the age of Darwin*. Chicago: University of Chicago Press.
Stavy, R., & Tirosh, D. (2000). *How students (mis-)understand science and mathematics: intuitive rules*. New York: Teachers College Press.
Tooby, J., & Cosmides, L. (2005). Conceptual foundations of evolutionary psychology. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 5–67). New York: Wiley.
Vinner, S. (1997). The pseudo-conceptual and the pseudo-analytical thought processes in mathematics learning. *Educational Studies in Mathematics*, *34*, 97–129.
Wilensky, U. (1991). Abstract meditations on the concrete and concrete implications for mathematics education. In I. Harel & S. Papert (Eds.), *Constructionism*.

Rethinking Probability Education: Perceptual Judgment as Epistemic Resource

Dor Abrahamson

Abstract The mathematics subject matter of probability is notoriously challenging, and in particular the content of random compound events. When students analyze experiments, they often omit to discern variations as distinct outcomes, e.g., HT and TH in the case of flipping a pair of coins, and thus infer erroneous predictions. Educators have addressed this conceptual difficulty by engaging students in actual experiments whose outcomes contradict the erroneous predictions. Yet whereas empirical activities per se are crucial for any probability design, because they introduce the pivotal contents of randomness, variance, sample size, and relations among them, empirical activities may not be the unique or best means for students to accept the logic of combinatorial analysis. Instead, learners may avail of their own pre-analytic perceptual judgments of the random generator itself so as to arrive at predictions that agree rather than conflict with mathematical analysis. I support this view first by detailing its philosophical, theoretical, and pedagogical foundations and then presenting empirical findings from a design-based research project. Twenty-eight students aged 9–11 participated in tutorial, task-based clinical interviews that utilized an innovative random generator. Their predictions were mathematically correct even though initially they did not discern variations. Students were then led to recognize the formal event space as a semiotic means of objectifying these presymbolic notions. I elaborate on the thesis via micro-ethnographic analysis of key episodes from a paradigmatic case study. Along the way, I explain the design-based research methodology, highlighting how it enables researchers to spin thwarted predictions into new theory of learning.

A few years ago, I was kicked out of a party. I had told the hostess that I study people's intuition for probability, and she wanted to hear more. So I did what I usually do in such situations: I asked her the two-kids riddle. Here's a rough transcription of our dialogue.

Dor: I have two kids. One of them is a girl. What's the sex of my other kid?

D. Abrahamson (✉)
University of California at Berkeley, 4649 Tolman Hall, Berkeley, CA 94720-1670, USA
e-mail: dor@berkeley.edu

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_13,

Hostess: I don't get it. It's just, like, 50–50—it's the same chance of getting a boy or a girl. What does your other kid have to do with it? They're totally independent events!

Dor: Indeed they *are* independent events, but together they make a compound event. There's double the chance that my other kid is a boy than a girl.

Hostess: Well excuse me, but that just doesn't make any sense at all. How could this possibly be true? *And* I resent your patronizing tone.

Dor: You see, two-kid families are either Girl–Girl, Girl–Boy, Boy–Girl, or Boy–Boy. That's all the possibilities. Now, we know that one of my kids is a girl, so that rules out Boy–Boy. So we're left with Girl–Girl, Girl–Boy, Boy–Girl. This means there's only one option with the other kid being a girl, but there're two options with the other kid being a boy. So it's 2:1. Boy wins.

Hostess: No you don't. That's just a mathematical model. In reality, it's still 50–50.

She became very upset and things got out of hand. She said that mathematicians should not pry their logical tentacles into that which is most sacred—the life of an unborn child. This was Berkeley, so I took it all with a grain of jalapeño. I went home and wrote a paper—by all accounts a favorable outcome for an untenured professor.

Still, probability has a bad reputation as being notoriously unintuitive—its assumptions opaque, its solution procedures arbitrary (see in Prediger 2008; Rubel 2009; Vos Savant 1996). So much so, that Kahneman et al. (1982) suggest we should refrain from applying gut feelings in making decisions with respect to complex randomness situations, and Fischbein (1987) advises substituting error-prone "primary intuition" with mathematically informed "secondary intuition."

Is this to be our fate? Is randomness to be indefinitely outside the province of informal reasoning? Are stochastic phenomena beyond the reach of our limited hominid brains? Perhaps textbook chapters on probability should begin with the banner, "Abandon intuition, all ye who enter here."

Perhaps, rather, the issue is not so much about probability phenomena or content per se but with how we teach this mathematical topic. Perhaps we choose to abandon intuition, because prevalent pedagogical praxis structurates probability content as patently alien to our gut feelings. Perhaps we all have remarkable informal intuitions for probability, only that traditional curriculum does not enable us to make use of this intuition beyond the very obvious cases of a single coin, die, or spinner. Perhaps there are better ways of introducing learners to the more advanced ideas of probability. Perhaps there are alternative materials and activities that empower students to engage their informal reasoning so as to build trust with the basic notions of the formal analytic frameworks for probability. Perhaps, specifically, informal intuition could be leveraged in the more complex cases of *two* coins, dice, spinners. At the same time, perhaps we would need to design a new compound-event random generator that is mathematically analogous to a pair of coins but is better tailored to accommodate relevant perceptual capacity.

Granted, informal intuition stemming from tacit perceptual capacity is a curious epistemic resource. It may not be as accurate as stipulated by the disciplinary context, in that it yields qualitative heuristic estimates where precise quantitative indices are required. Moreover, intuition is susceptible to context, in that its successful application may require very particular forms of sensory presentation and, as such, does not transfer to semiotic systems typical of professional practice, such as alphanumerical symbols. Finally, intuitive reasoning may not lend itself to reflecting on the reasoning process and documenting it in forms that enable scrutiny of peers.

That said, intuitive reasoning can ground formal knowledge by evoking meanings, schemas, familiarity, and coherence that can be brought to bear as we approach unfamiliar mathematical analyses, tools, procedures, and displays. Furthermore, intuitive reasoning enables instructors and students to share referents—they can talk *about* properties of phenomena, such as chance events, even before these phenomena have been formalized. As such, informal reasoning can be a useful epistemic resource for learning formal models in discursive contexts. Yet for this to work, informal inference needs to align with mathematical analysis. Enter designers.

In this chapter, I propose a rethinking of how students should be introduced to the fundamental principle of classicist probability theory, the rule of ratio, as it applies beyond the simple cases of single events. I propose that students' perceptual judgment of the stochastic propensities inherent to random generators should constitute an epistemic resource for making sense of the classicist approach to probability, particularly in the case of compound events. I argue that *under auspicious conditions, students' perceptual judgment of the stochastic propensities of a random generator can play a similar epistemic role as do actual experiments with the device in terms of evoking sensations of relative likelihood that, in turn, can be linked to the distribution of possible outcomes in the event space*. In order to create these auspicious pedagogical conditions, I submit, educational designers should create materials and activities geared to accommodate humans' evolved perceptual inclinations, such as sensitivity to proportional relations in the visual field. The chapter attempts to promote this rethinking by furnishing intuitive, empirical, and theoretical evidence as support for its validity.

The conjecture that probability theory can be grounded in perceptual judgment drove a multi-year design-based research project. I will describe how the design was rationalized, engineered, and ultimately implemented over a set of iterated studies. In particular, the conjecture became refined through this designer's reflection on his own intuitive rationales for the particular materials, activities, and facilitation developed over the project (Schön 1983; Vagle 2010). These reflections emanated from analyses of empirical data from implementing the design, specifically video footage from clinical–tutorial interactions, in which 28 9-to-11 year old children and 24 undergraduate/graduate statistics students engaged the activities individually.

The more general form of my conjecture, namely that conceptual understanding should be grounded in perceptual reasoning, is hardly new to the learning sciences literature. I will draw on seminal theoretical frameworks to suggest that *unmediated, pre-analytic perceptual sensation from a phenomenon under scrutiny may be an equally or even more powerful epistemic grounding for a mathematical model of the phenomenon than mediated, encoded information about the phenomenon's behaviors*. For example, visually analyzing a random compound-event generator may better ground its event space than visually analyzing its actual experimental outcome distribution. This suggestion invites a potentially interesting dialogue with those who claim that students' inferences should be encoded directly in formal symbolic systems (Cheng 2011). The suggestion also implies a central role for designers to customize phenomena so as to render them cognitively congenial (Kirsh 1996) or cognitively ergonomic (Artigue 2002), and specifically more conducive to the application of tacit perceptual capacity relevant to the targeted mathematical concepts (Abrahamson 2006a; Abrahamson and Wilensky 2007).

I begin, below, by focusing on students' difficulty with the logic of permutations, for example listing only [Heads–Tails] as the mixed result of a two-coin experiment rather than [Heads–Tails *and* Tails–Heads]. The conjecture evaluated in this paper was originally motivated as a response to students' difficulty with this particular situation.

1 Compound Event Spaces: Rethinking an Enduring Pedagogical Challenge

> Adaptors [of mathematics to the school level] do not trust their eyes, they cannot believe it is so simple, or if they can, they do not trust other people to be able to believe it. (Freudenthal 1974, p. 227)

In 1814, Pierre-Simon Laplace put forth a theorem, by which the likelihood of an event occurring randomly is the ratio of the total number of outcomes *favorable* to the occurrence of that event and the total number of *possible* outcomes, where both uniqueness and equiprobability of outcomes are assumed. Implicit to a successful application of this algorithm is the construction of an event space that includes all possible, unique outcomes. Here lies the educational rub.

1.1 The Problem of Differentiating Outcomes in a Compound Event Space

Consider a fair coin tossed onto a desk. Students learning probability by-and-large readily accept that the chance of receiving a Heads is one-in-two. Formally, "one" is the total number of favorable outcomes (we discern only one way of getting Heads), and "two" is the total number of all possible outcomes (we discern exactly two possible outcomes: Heads, Tails). By and large, students familiar with the notion of proportions rarely encounter substantial difficulty in constructing a simple event space and applying the Laplace principle.

Now consider a pair of coins. What is the chance of receiving Heads-and-Tails? The event "Heads-and-Tails" is compound, meaning that its occurrence depends on the intersecting occurrences of (two) simple events, here a Heads *and* a Tails. A pair of coins has a total of four possible outcomes. This number can be calculated by multiplying the total number of possible outcomes in each coin, so $2 \times 2 = 4$. These four possible outcomes are [HH, HT, TH, TT], and two of them, HT and TH, are favorable to the occurrence of the compound event in question Heads-and-Tails, whose chance is thus $2/4$, that is, $1/2$.

An enduring educational challenge lies in having students list a complete event space for compound-event experiments built of *identical* random generators, such as two pennies. Students often do not distinguish between variations, that is, possible outcomes that differ only in the order of constituent singleton events, such as HT and TH. Of course, students can *see* that "HT" and "TH"—the actual Roman characters written on paper—are different symbol strings, and they can learn to *enact* the combinatorial analysis algorithm by which such variations are generated. The issue is not visual or procedural per se but logical, conceptual. That is, in the context of analyzing a random generator, students typically claim that variations are redundant because the variations do not indicate different worldly eventualities; they believe that only one of these variations should be listed in the event space. For example, students build for the two-coin experiment an event space of only three possible outcomes—[HH, HT, TT]—and consequently determine via the rule of ratio a $1/3$ chance of receiving "Heads-and-Tails" (Abrahamson and Wilensky 2005b). This form of reasoning, which does not agree with probability theory or empirical results, is considered an impediment to the teaching and learning of probability. The ubiquity and perseverance of this reasoning throughout the school years and into college makes it a major pedagogical challenge (Batanero et al. 1997; Fischbein and Schnarch 1997).

Educators have sought to help students make sense of the Laplace principle as it applies in the case of compound events. In particular, designers and teachers have attempted to support students in understanding *that* and, ideally, *why* listing all variations on an event, such as both HT and TH, is critical to building the event space of a compound-event random generator. However, success has been mixed (Jones et al. 2007; Shaughnessy 1977).

A prevalent design rationale for introducing the analysis of compound events is to have students witness its predictive utility. That is, it is hoped that students accept the importance of including permutations in the event space of a compound-event random generator via engaging in activities wherein event spaces that include variations turn out better to predict its actual experimental outcome distributions than do event spaces that do not include variations. Below, we first elaborate on this design rationale and then critique it and offer an alternative design rationale and solution.

1.2 Empirical Experimentation as Epistemic Grounding for Classicist Analysis

Many education researchers believe that students best learn classicist theory when combinatorial analysis activities are combined with empirical activities (Steinbring

1991). In practice, students operate a random generator and reflect on the results vis-à-vis their expectations (Amit and Jan 2007). Some researchers use computer-based media to build microworlds wherein learners can author, run, analyze, and modify stochastic experiments. These chance simulators typically include animations of actual or hypothetical random generators as well as a variety of standard forms, such as graphs, tables, and monitors, that display cumulative results in real time (e.g., Abrahamson 2006b; Iversen and Nilsson 2007; Konold 2004; Pratt 2000; Stohl and Tarr 2002; Wilensky 1995).[1]

The underlying assumption of this empiricist approach is that learners are fundamentally rational beings. Learner are expected to accept experimental results as a valid epistemic resource bearing on the problem; accordingly, in the face of results that patently contradict their expectations, they will acknowledge the inadequacy of their reasoning and seek an alternative form of reasoning that better accounts for these results; in particular, they will recognize that compound event spaces *with* variations are better fit to the world and thus should be preferred over spaces without permutations (cf. Karmiloff-Smith 1988; Koschmann et al. 1998; von Glasersfeld 1987). For example, students whose erroneous analysis led them to expect that only a $1/3$ of the two-coins experimental outcomes will be of the type Heads-and-Tails may take pause when simulations of this experiment repeatedly converge toward $1/2$ (Wilensky 1995). In sum, students are expected to endorse experimental outcomes as veridical and accordingly seek to adjust their reasoning that has proven erroneous.

Being proven wrong is a formative learning experience, because it stimulates reflection, which may result in conceptual change. As such, experimental empiricism appears to provide an epistemic resource that promotes student reasoning in ways that agree with the overall pedagogical objective.

Yet being proven wrong might bear also implicit effects that transcend the content itself and result in pedagogically undesirable habits of mind (see Bateson 1972, on deutero learning). That is, being proven wrong may cause students to distrust their intuitive judgment as an epistemic resource and develop epistemological anxiety with respect to the formal procedure (Wilensky 1997); students may learn to rely solely on empiricism, which lends a sense of certainty but not a deeper sense of causality (Harel 2013); students will consequently forget a procedure they studied that is "not tied to lived reality with strong bonds" (Freudenthal 1971, p. 420). Given these potentially deleterious effects of being proven wrong, are alternative pedagogical approaches feasible?

Perhaps not. Some researchers believe that humans' primary intuitions for compound-event probabilistic situations are inherently fallible (Kahneman et al.

[1] Some of this design work bears the concomitant far-reaching conjecture that cyber media can and should *transform* mathematics—not only serve mathematics as we know it—by offering alternative cognitive access to phenomena in question (Wilensky and Papert 2010). For example, as human practice becomes increasingly "wired," brute-force frequentist simulation could render classicist analysis a redundant anachronism shelved away with other precyber artifacts and procedures, such as logarithmic tables, slide rules, and long division.

1982). It follows that humans must perforce resign themselves to developing secondary intuitions better fitted to rational models of stochastics (Fischbein 1987).

However, it could be that humans' primary probabilistic intuition is in fact well aligned with mathematical theory; studies that concluded otherwise may have been inauspicious to primary intuition by using materials and tasks that did not accommodate perceptual reasoning (Gigerenzer 1998; Xu and Garcia 2008; Zhu and Gigerenzer 2006). More emphatically, I maintain that naïve reasoning, no matter how incompatible it might seem with respect to formal knowledge and practice, should be conceptualized as a resource—not an impediment—to designing, teaching, and learning (Borovcnik and Bentz 1991; Bruner 1960; Gigerenzer and Brighton 2009; Smith et al. 1993; Wilensky 1997). More broadly, if somewhat lyrically, mathematics is a human invention—it was created in the image of (wo)man—and so other humans are biologically and cognitively equipped to share these images (Núñez et al. 1999).

I have thus furnished theoretical, pedagogical, and philosophical critiques of the prevalent approach to introducing the mathematical content of compound event spaces. These critiques suggest that empirical experimental evidence might not be the *sine qua non* epistemic resource for students to ground the logic of compound event spaces. I wish to propose perceptual judgment as an epistemic resource alternative, or at least complementary, to experimental empiricism. Given appropriate design, I maintain, perceptual judgment may enable learners to ground compound event spaces in correct rather than incorrect naïve judgment and thus relate to this content, beginning with its introduction and even on beyond.

1.3 Perceptual Judgment of Likelihood as Epistemic Grounding for Classicist Analysis

Even before one analyzes a random generator so as to build its event space, one is sometimes able to offer estimates for its stochastic propensities. In order for these naïve judgments to ground the event space, though, they must align with probability theory. Lo, have we not been arguing that students' naïve probabilistic reasoning about compound-event random generators is the very source of their difficulty to accept the combinatorial analysis procedure?

Let us interrogate the context inherent to that argument. In particular, I will attend to the particular types of *materials* and *tasks* related to that argument, in an attempt to understand how these dimensions affect the compatibility of informal and formal probabilistic reasoning. I begin with the task.

When earlier we discussed an apparent incompatibility between students' naive reasoning and formal probability theory, the task that students were erring on was combinatorial *analysis* of a compound-event experiment. Still, students' *pre-analytic* reasoning, and in particular their visual judgments about the random generator's stochastic propensities, may be compatible with formal probability theory. If so, then perceptual judgment might constitute an epistemic resource for grounding formal treatment of these experiments. The formal event space thus would not conflict with perceptual judgment. Rather, the event space would triangulate, explain, or elaborate perceptual judgment.

The materials of the learning task—that is, the particular random generator—are also of relevance. In our earlier expository discussion, the compound-event random generator was a pair of coins. Perhaps students' erroneous analyses of this particular device are due to it being non-conducive to pre-analytic perceptual reasoning aligned with mathematical theory. Perhaps there are alternative devices that are conducive to humans' primitive statistical inclinations. If so, perceptual judgment of these devices' propensities might complement or even improve on frequentist experimentation as epistemic resources for grounding compound event spaces.

The next section of this chapter describes a design-based research project that explored perceptual judgment as an epistemic resource for grounding combinatorial analysis of compound event spaces. Coming into the study cycles, I asked: *What structural forms might a random generator take that would elicit from students perceptual judgments in agreement with mathematical theory of probability?* Later in the cycles, when I had a working answer for this design question, a new question emerged: *Once we elicit from students perceptual judgments that agree with mathematical theory, how might the students come to express these judgments in mathematical form?*

These questions are important to mathematics education theory and practice. Findings from a research program pursuing these questions may imply that dice, coins, spinners, and slot machines—ecologically authentic random generators that played formative historical roles in the development of probability theory—do not make for optimal *introductory* materials for this mathematical subject matter, because they do not afford pre-analytic epistemic resources aligned with classicist theory as it obtains in the case of compound events.

Up to now, I have discussed an *intuitive* rationale for the proposed "perceptual approach" to grounding probability. Section 1.4, below, bridges toward the *empirical* part of the chapter by explaining how theory evolves in design-based research projects. Section 2 then describes a design-based research project that investigated probabilistic reasoning. Interleaved into the design narrative is a Piagetian–Vygotskian framework that evolved through the course of analyzing empirical data gathered over that project. The framework underscores the role of pre-analytic perceptual judgment in learning mathematical content. Section 3 draws on *theoretical* resources from the learning sciences to further support the framework. Section 4 integrates these intuitive, empirical, and theoretical supports for the chapter's central conjecture and offers conclusions.

1.4 The Evolution of Theory Through Cycles of Design-Based Research Studies

The conjecture pursued in this chapter is that students may bear pre-analytic cognitive resources that are suitable for grounding fundamental notions of probability theory, only that curricular materials and tasks commonly employed in introductory probability units do not enable students to draw on these resources; by understanding the nature and function of these subjective resources, we may be able to engineer tasks and materials that better accommodate and leverage these resources. This chapter describes a research project that put the conjecture to the empirical test (see, e.g., Confrey 2005, on the design-based research approach).

The rethinking of probability pedagogy proposed in this chapter evolved through design-based research in the domain of probability, and the assertions offered here were originally couched in terms of particular materials and activities developed in that project. More specifically, the rethinking is the researcher's articulated reconceptualization of his own design in light of multi-year analyses of empirical data gathered in its iterated implementations.

I will support the rethinking via tracing its evolution along project milestones (Collins 1992). Each milestone is an instance where, having wrestled with an enduring challenge of analyzing empirical data gathered in our studies, I realized the purchase of a hitherto unconsidered theoretical model and methodology, typically from the learning sciences literature. Yet even as the research team embraced these theoretical models as tools applied to the empirical data, the models' evident analytical utility impelled us to reconsider our theoretical assumptions. Namely, when multiple models are all taken to bear on the same data, one can promote theory via attempting to resolve tension among the models' epistemological underpinnings (Artigue et al. 2009; Cobb and Bauersfeld 1995; Sfard and McClain 2002).

Indeed, I will offer a view on probability education that integrates cognitivist, sociocultural, and semiotics frameworks. I will propose a theorization of mathematical learning as guided, heuristic–semiotic coordination of sensations from two different types of epistemic resources: (a) tacit gestalts rooted in perceptual primitives; and (b) formal analysis mediated by artifacts.

From a constructivist perspective, this view of mathematics learning implies a pedagogical dilemma that revisits Plato's *Meno*: (i) on the one hand, students require gestalt perceptions of the phenomena they are studying—they need to know what they are learning *about*—even before they engage in analyzing it (Bereiter 1985; Pascual-Leone 1996); on the other hand, (ii) these gestalt meanings ironically cause students to resist formal analysis of these phenomena, because the analysis carves the phenomena along dimensions that appear to the child irrelevant to the intuitive gestalt perception (Bamberger and diSessa 2003). In particular, how might students ground "order-based" compound-event spaces (e.g., [HH HT TH TT]) in "orderless" gestalt sensation of event representativeness or anticipated plurality?

From sociocultural perspectives, this incompatibility does not present a dilemma, because learning in the disciplines necessarily involves conceptual reorganization (Newman et al. 1989). Notwithstanding, I maintain, mathematical models of situated phenomena require pre-analytic gestalts as their *aboutness*, that is, as their core epistemic resource.

The following section elaborates on this conceptualization of learning in the context of the design-based research project wherein the conceptualization evolved.

2 Pursuing a Design-Based Research Conjecture to Build Learning Science Theory

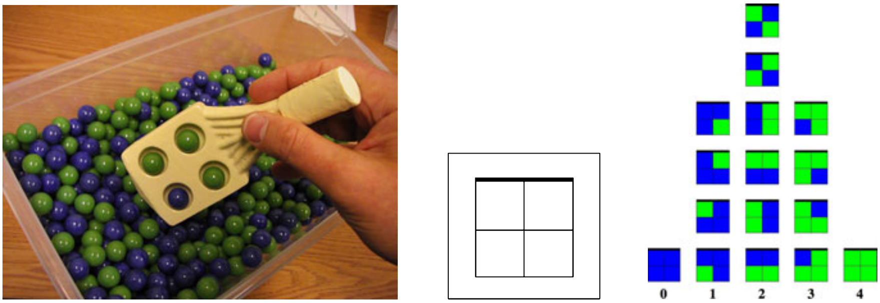
This chapter argues for the epistemic role of perceptual judgment in grounding probability notions. The thesis evolved through a set of empirical studies carried out over a multi-year educational research project that investigated issues of teaching, learning, and design pertaining to probabilistic reasoning. The project, initially named *ProbLab*, began during my Postdoctoral Fellowship at the Center for Connected Learning and Computer-Based Modeling at Northwestern University (Wilensky, *Director*) and continued as *Seeing Chance*, a NAE/Spencer Postdoctoral Fellowship at the Embodied Design Research Laboratory at the University of California, Berkeley, which I direct.

The design-based research project drew inspiration from the *connected probability* work (Wilensky, 1993, 1995, 1997). The project grew into a three-pronged effort to develop products, theory, and frameworks (see Abrahamson and Wilensky 2007, p. 25). Embarking from prior findings (Abrahamson and Wilensky 2002, 2004, 2005a, 2005b), I: (a) built mixed-media materials and activities for learning fundamental probability notions (Abrahamson 2006b) and evaluated their pedagogical affordances (Abrahamson 2007; Abrahamson and Cendak 2006); (b) developed explanatory models for the roles of perceptual reasoning in conceptual learning (Abrahamson 2009b, 2010, 2011, 2012b; Abrahamson et al. 2012); and (c) created a grounded mathematics design framework (Abrahamson 2009a) and contributed to reflective discourse on design-based research practice (Abrahamson 2009c, 2012a).

The project was formative in shaping a theoretically balanced perspective on mathematics learning. The perspective emerged gradually through cycles of analyzing video footage gathered during the *Seeing Chance* tutorial interactions ($n = 28$, Grades 4–6; $n = 24$, college seniors and graduate students). The collaborative analyses focused on episodes within these data that culminated in students expressing meaningful links among resources in the learning environment in ways that we evaluated as pedagogically desirable, in that they promoted the didactical objective of the interaction. Our theory development was motivated by a sense that the analytic tools we were using were not affording productive interpretations of these data episodes. As such, the theory development process was contingent on the researchers acknowledging the limitations of their analytic tools and, more deeply, problematizing the theoretical assumptions tacitly informing the selection of those tools.

At it broadest, I wanted to understand how people articulate tacit judgment in mathematical form. I therefore showed students perceptual displays bearing quantitative relations, asked them a framing question related to those properties, and then introduced semiotic resources for them to express their inferences. In creating the perceptual displays, I sought to enable students to infer qualitative judgments that agree with mathematical theory.

Specifically for the subject of probability, I created a set of recourses (see Fig. 1): (a) a concrete random generator; (b) media for building an event space via combinatorial analysis of the random generator; and (c) an innovative structural form for



(a) (b) (c)

Fig. 1 Materials used in a design-based research project investigating relations between informal intuitions for likelihood and formal principles of an event space: (**a**) a "marbles scooper," a utensil for drawing out ordered samples from a box full of marbles of two colors; (**b**) a card for constructing the sample space of the marbles-scooping experiment (a stack of such cards is provided, as well as a green crayon and a blue crayon, and students color in all possible outcomes); and (**c**) a "combinations tower," a distributed event space of the marbles-scooping experiment, structured so as to anticipate the conventional histogram representation of actual outcome distributions

organizing the event space so as to make it more conducive to heuristic perceptual inference. In addition, I designed and built a suite of computer-based simulations of the experiment featuring schematic models of the random generator. In this chapter, I do not feature these computer-based simulations, because my thesis here pertains primarily to students' perceptual judgments of the random generator itself and, in particular, how students coordinated these judgments with their guided perceptions of the event space. But I will briefly mention the simulations in the general discussion so as to compare and contrast perceptual judgments and empirical experimentation with respect to their epistemic contributions to content learning.[2]

I first had students briefly examine the random generator's experimental mechanism (see Fig. 1(a)). Immediately after, I asked them to offer their estimations for its expected outcome distribution. Importantly, *the students did not conduct any actual experiment at all*, so that their responses were based only on sensory perception and reasoning.

The particular random generator designed for this project was such that students tended to offer likelihood judgments that agreed with mathematical theory. In particular, they predicted a plurality of two-green-and-two-blue samples (hence, "2g2b"), a rarity of both 4b and 4g, and, in between, the events 1g3b and 3g1b. When asked to support their responses, students referred to the equal number of green and blue marbles in the bin.

[2] Strictly speaking, the marbles-scooping experiment is hypergeometric, not binomial, because as each marble is captured by a concavity in the scooper, there is one less of that color in the bin. However, the fairly minute ratio of the sample (4) to the total number of marbles in the bin (hundreds) enables us to think of this experiment as quasi-binomial and, for all practical effects, as actually binomial.

Next I guided students to build the experiment's event space. Instead of having them represent the possible outcomes as a list or tree of symbols on a single sheet of paper, I provided a stack of cards (see Fig. 1(b)) as well as two crayons, and students used these media to create iconic representations of the possible outcomes. Typically, students organized the construction space on the desk by clustering the cards into five emergent groups, with 1, 4, 6, 4, and 1 items, respectively (Abrahamson 2008). I then guided the students to assemble the 16 cards according to these five event classes in a spatial configuration that highlighted the different number of outcomes per each event (see Fig. 1(c)).

In line with my theoretical stance coming into the study, I had conceptualized the event space as a formal concretization of intuitive reasoning. I therefore expected that the event space would stimulate students to abstract and encapsulate the schemas they had tacitly employed in judging the likelihoods of the device's possible outcomes (Piaget 1968). Specifically, I expected the students to recognize the variations—such as [gggb, ggbg, gbgg, bggg]—as articulating their pre-analytic judgments. In so doing, I envisioned and hoped for a smooth continuity from naïve to informed views of the experiment; I did not envision tension between these views.

My expectation was largely informed by the great care I had taken in the design to build a sampling device (the scooper) that configured each sample in a particular order. That is, I had designed "order" as an inherent structural property of the sampling device, because I had thought that this feature would impress upon students the uniqueness of the variations. I therefore implicitly took it for granted that the students were attending to the dimension of order. It never occurred to me that they might see the scoops different from how I saw the scoops; that they could construe gggb and ggbg as "the same thing."

However, throughout the construction of the event space, students tended to resist the variations as redundant objects. They thought that the event space should consist of five cards only, with one card per each of the five events 4b, 1g3b, 2g2b, 3g1b, 4g. They accepted the complete event space only once the combinations tower had been assembled. Specifically, *students endorsed the variations only once they were able to perceive the event space's five vertical projections as respectively signifying their sensation of the five events' differential likelihoods per their earlier perceptual judgments of the marbles box*. Only after having thus made sense of the event space in its totality as mapping onto their informal inference did the students retroactively accept the analytic procedure by which the event space had been constructed.

As such, my implicit assumption that tacit judgment can be directly articulated in disciplinary form quickly became problematized. In particular, my research team came to acknowledge our implicit assumption and recognize that it apparently built on a lopsided constructivist conceptualization of learning. We thus gradually began to realize the overwhelming constitutive role of artifacts in mediating cultural forms of reasoning (Wertsch 1985) as well as instructors' formative role in framing and guiding this mediation process (Newman et al. 1989).

Prior to this sociocultural calibration of our design rationale, we had focused exclusively on the student as the locus of education. We had regarded the phenomenon of learning essentially as the child's solipsistic developmental process, which the researcher merely catalyzed for the purposes of the study. As such, we had construed the experimenter's actions as necessary methodological means of eliciting from the study participant a targeted response. Yet the experimenter's procedural action, we came to realize, was in fact simulating a culturally authentic interaction between a tutor and student. In particular, we now surmised, *learners do not articulate tacit judgment directly in mathematical form; rather, they objectify these presymbolic notions in semiotic means made available to them in the learning environment* (Radford 2003). Moreover, student expression cannot always be direct articulation of the tacit judgment, because the available cultural media often require a parsing of the source phenomenon in ways that are very different from naïve perception (Bamberger and Schön 1983; Bamberger and diSessa 2003).

Still, if our study participants were struggling to objectify presymbolic notions in cultural forms, what was the source of these presymbolic notions? How did the students perform perceptual judgments in the first place, during the initial phase of the experiment, when they essentially gazed at the experimental utensils? Research on infants' "statistical" reasoning, in empirical settings strikingly analogous to ours (Xu and Garcia 2008), suggests that our study participants drew on early, unmediated perceptual capacity to judge the relative chances of random events. It could be that the order of singleton outcomes is ignored or downplayed in these judgments and only the color ratios are perceived.

Students thus appear to be working with two resources, tacit reasoning and cultural artifacts. *When tacit inference is aligned with mathematical theory, instructors can guide students to appropriate the cultural resource as a means of supporting and empowering their tacit inference*. This insight into design bears concomitant insight into theory of learning, as follows.

The sociocultural perspective, which underscores the mediating role of cultural artifacts in conceptual development, appears to require a constructivist complement so as together to afford a comprehensive explanation of what students are doing when they link tacit and cultural resources. We were thus heartened when diSessa (2008) called explicitly for "dialectical" theoretical work that seeks to combine and possibly synergize cognitivist and sociocultural models of learning. The call resonated with other urges to view the work of Piaget and Vygotsky as compatible (Cole and Wertsch 1996; Fuson 2009; Stetsenko 2002).

Once we had aligned our theorizing with the sociocultural perspective, we could infuse into the swell of our data analysis seminal neo-Vygotskian contributions to the modeling of mathematics teaching and learning (Bartolini Bussi and Mariotti 2008; Saxe 2004; Sfard 2002, 2007). Consequently, *we came to view mathematics learning as the elicitation and acculturation of subjective meanings via guided, goal-oriented, and artifact-mediated activity*. As a result of this dialectical theorizing, we could offer a sociocultural interpretation of abductive inferential reasoning (Abrahamson 2012b) as well as a reconceiving of discovery-based learning (Abrahamson 2012a).

This section answered our chapter's research questions regarding the prospects, materials, tasks, and process of grounding a compound event space in perceptual judgments of random generators. Having outlined the evolution of our perspective on mathematics learning, what remains is to reflect on the epistemic nature of perceptual judgment and then compare perceptual judgment to experimental empiricism as complementary epistemic resources for grounding compound event spaces.

3 Discussion: Learning Sciences Views on Perceptual Judgment as Epistemic Resource

What is the role of perceptual judgment in mathematical learning? Is it desirable for students to reason about properties of concrete objects, given that mathematical texts are symbolic? What are the epistemological and cognitive qualifications of a proposal that perceptual judgment of random generators can serve as an epistemic resource for understanding compound event spaces?

Drawing on a broad reading of the learning sciences literature, I will now build the argument that perceptual judgments, and more generally multimodal dynamical images, are essential for conceptual development—they are what mathematical texts are *about*.

The human species developed via natural selection the capacity to quantify aspects of phenomena relevant to their survival (Gigerenzer 1998). Some of these enabling constraints on perception (Gelman 1998) pertain to quantities whose mathematical modeling features in school curriculum, such as the intensive quantities of slope, velocity, chance, and aspect ratio (Suzuki and Cavanagh 1998; Xu and Garcia 2008). Do we, therefore, like the slave in *Meno*, know the concepts before studying them?

Not quite. Perceptual capacity cannot be directly translated into mathematical knowledge. First, perceptual judgments are holistic and pre-articulated, such as when we perceive the gradient of a sloped line, whereas mathematical modeling is analytic and symbolic, such as when we measure and calculate "rise over run." Second, perceptual judgments are tacit—the neural mechanisms of these cerebral faculties are cognitively impenetrable (Pylyshyn 1973), and so we are conscious not of our perceptual process itself but only of our operatory reactions and contextual inferences that result from these tacit processes. Referring to the relation between object constancy and proportional reasoning, Piaget and Inhelder (1969) wrote, "However elementary they may be in the child, these concepts cannot be elaborated without a logico-mathematical structuration that… goes beyond perception" (p. 49).

From a neo-Vygotskian perspective, this logico-mathematical structuration of perceptual judgment is achieved via social mediation. In particular, individuals appropriate cultural forms as means of realizing personal goals for solving collective problems (Saxe 2004). In interactive contexts, such as tutorial or classroom activities, these personal goals may be discursive. In particular, the semiotic–cultural perspective (Radford 2003) highlights the role of students' presymbolic notions in educational interaction: students develop new mathematical signs by objectifying presymbolic notions using available semiotic means.

Educators can thus play vital roles in students' conceptual development by strategically placing pedagogically desirable cultural forms in the learning environment and steering students to re-articulate their naïve views by these particular semiotic means (Abrahamson 2009a; Mariotti 2009; Sfard 2002, 2007). Students may appropriate cultural forms also as means of accomplishing enactive goals, and not just discursive goals, and in so doing they may bootstrap new operatory schemas by reconfiguring their naïve strategies (Abrahamson et al. 2011).

I thus discern across a range of constructivist and sociocultural theorists a loose consensus, by which meaningful learning occurs when individuals instrumentalize cultural forms to accomplish tasks (Vérillon and Rabardel 1995). When the tasks are embedded in pedagogical activities and the forms are mathematical symbolic artifacts, students become acculturated into mathematical discourse and praxis.

Implicit to this dialectical process by which personal sense meets cultural form is the question of the cognitive form of personal sense, prior to its acculturation. Many scholars believe that personal sense is embodied as multimodal images. For example, scholars in mathematics (Davis 1993), semiotics (Hoffmann 2003; Peirce 1867), developmental psychology (Arnheim 1969), creativity (Getzels and Csikszentmihalyi 1976; Hadamard 1945), and philosophy (Barwise and Etchemendy 1991) believe that perceptual reasoning, such as visual or imagistic pattern recognition, is the *sine qua non* of conceptual development. Moreover, according to theories of cognition (Barsalou 1999; Glenberg and Robertson 1999) and cognitive linguistics (Goldin 1987; Johnson 1999; Lakoff and Johnson 1980; Lakoff and Núñez 2000), all reasoning is perforce imagistic by virtue of the fundamental cerebral architecture and mechanisms of reasoning and perception. It is therefore hardly provocative to explore mathematical pedagogy that fosters perceptual grounding for symbolic text (Kamii and DeClark 1985).

4 Conclusion: Perceptual Judgment Grounds Classicist Analysis

Perceptual judgment of random generators enables students to meaningfully ground the products of formal classicist analysis procedures. As such, perceptual reasoning constitutes a viable pedagogical entry into fundamental probability content, and particularly into compound event spaces.

I began this chapter by illustrating the need for effective probability curriculum. I then underscored the importance of designing materials and tasks appropriate to leveraging the epistemic resource of perceptual reasoning. Next, I demonstrated the plausibility of my thesis via describing milestones in a decade-long empirical design-based investigation of probability learning. Finally, I supported the conjecture with seminal theory from the learning sciences literature.

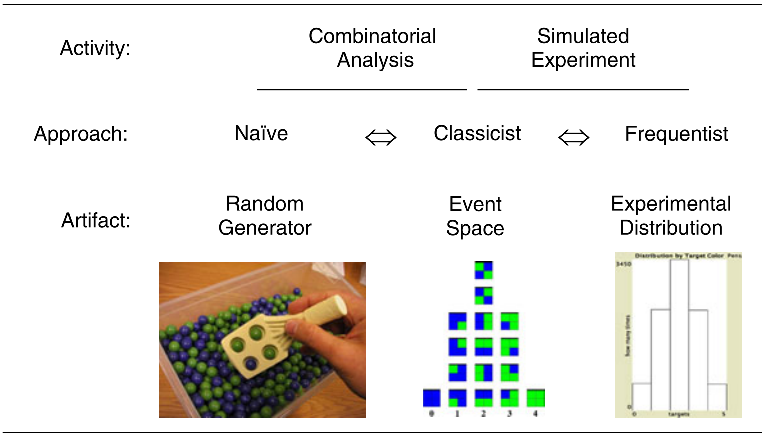


Fig. 2 Probability design trialogue. Conceptually critical coordination via two activities across three artifacts in a design for the binomial. Activities bridge complementary conceptualizations of the stochastic phenomenon. *Double arrows* indicate that learners need to interpret a new artifact they encounter as signifying meanings they had established for a previous artifact. The original device suggests its own stochastic propensity, the event space models the propensity, and the experimental distribution exemplifies the propensity. Both naïve and frequentist conceptualizations ground the classicist artifact. Accepting the event space retroactively grounds the combinatorial analysis procedure by which the space was built

The conjecture that this chapter has sought to promote should not by any means discourage educators from employing frequentist approaches in the instruction of probability. *A fortiori*, the empirical data discussed in this chapter represented only two of my three design phases, where the third phase consisted of running computer-based simulations of the stochastic experiment. In those activities, I guided students to draw on their intuitive sensations both from the static random generator and its event space so as to make sense of actual, "imperfect" outcome distributions that resulted from the experimental runs (Abrahamson 2007, 2010). Linking intuitive, analytic, and empirical probabilistic activities appears to support a coherent and connected perspective on probability (Wilensky 1993).

Perceptual judgment of random generators and empirical experimentation with random generators play different *curricular* roles in terms of the conceptual content they explore. For example, perceptual examination of a random generator is a condition for its combinatorial analysis, whereas conducting experiments with the random generator creates opportunities to encounter randomness and sample size as they relate to variance. However, perceptual judgment and empirical experimentation play similar *epistemic* roles in understanding event spaces (see Fig. 2): both activities evoke sensations that resonate with the distribution of possible outcomes across events; both activities may result in implicating the event space as explaining the random generator's propensities that we sense or witness; in both cases, adopting the event space is mediated by tacit or direct sensation of relative magnitudes.[3]

Notwithstanding, naïve perceptual judgment of random generators is different from experimental activity, in that perceptual judgment directly evokes presymbolic notions of the property in question, whereas experimental outcome distributions indirectly evoke these notions. This difference between immediate and mediated notions may confer uper perceptual judgment a unique advantage over experiments in relation to selecting introductory grounding activities for probability designs. Elsewhere, we have demonstrated the extensibility of these introductory activities toward incorporating symbolic displays as well as cases of heteroprobable outcomes (Abrahamson 2009a).

Students possess natural capacity to perform powerful perceptual reasoning pertaining to the study of probability. Designers, teachers, and researchers may greatly avail themselves by leveraging this power so as to support the learning and continued investigation of this chronically challenging subject matter.

Acknowledgements I wish to dedicate this chapter to my dear friend Ólafur Elíasson, whose noble recording of the Adagio from J.S. Bach's piano concerto in F minor inspired me as I wrote. Thanks to Maria Droujkova for excellent formative comments on an earlier draft. I really do have two kids: a girl and a boy (in that order).

References

Abrahamson, D. (2006a). Mathematical representations as conceptual composites: implications for design. In S. Alatorre, J. L. Cortina, M. Sáiz, & A. Méndez (Eds.), *Proceedings of the twenty eighth annual meeting of the North American chapter of the international group for the psychology of mathematics education* (Vol. 2, pp. 464–466). Mérida: Universidad Pedagógica Nacional.

Abrahamson, D. (2006b). The shape of things to come: the computational pictograph as a bridge from combinatorial space to outcome distribution. *International Journal of Computers for Mathematical Learning*, *11*(1), 137–146.

Abrahamson, D. (2007). *The real world as a trick question: undergraduate statistics majors' construction-based modeling of probability*. Paper presented at the annual meeting of the American Education Research Association, Chicago. http://gse.berkeley.edu/faculty/DAbrahamson/publications/Abrahamson-AERA2007.pdf.

[3]The term *trialogue* in Fig. 3 is borrowed from Wilensky (1996) who, shifting attention to artifacts rather than activities, writes, "By engaging in computational modeling—this trialogue between the symbolism, the program output and the real world—and, then, reflecting on the feedback obtained, learners can make meaningful connections" (p. 128). Wilensky's assertion was expressed in the context of constructionist activities, wherein learners themselves create the computer-based models, and so the "symbolism" element of the trialogue refers to alphanumerical expressions in the modeling language (the "code," such as "forward 10"). Nevertheless, Wilensky's notion of a trialogue among inquiry artifacts obtains in the case of ready-made models, too, such as in the case of *Seeing Chance* design discussed in this chapter. Therein the "symbolism" element refers to inscriptions generated via analyzing the real world, such as diagrams, icons, and concrete displays (i.e., the "combinations tower" event space).

Abrahamson, D. (2008). Fostering the emergence of an embodied cognitive artifact: the case of the number line in a design for probability. In D. Abrahamson (Chair), D. Earnest (Org.), & H. Bass (Discussant), *The many values o the number line—an interdisciplinary forum. Symposium presented at the annual conference of the American Education Research Association*, New York, March 24–28. http://gse.berkeley.edu/faculty/DAbrahamson/publications/Abrahamson-AERA2008-NumberLine-symp.pdf.

Abrahamson, D. (2009a). Embodied design: constructing means for constructing meaning. *Educational Studies in Mathematics*, *70*(1), 27–47. Electronic supplementary material at http://edrl.berkeley.edu/publications/journals/ESM/Abrahamson-ESM/.

Abrahamson, D. (2009b). Orchestrating semiotic leaps from tacit to cultural quantitative reasoning—the case of anticipating experimental outcomes of a quasi-binomial random generator. *Cognition and Instruction*, *27*(3), 175–224.

Abrahamson, D. (2009c). A student's synthesis of tacit and mathematical knowledge as a researcher's lens on bridging learning theory. In M. Borovcnik & R. Kapadia (Eds.), Research and developments in probability education [Special Issue]. *International Electronic Journal of Mathematics Education*, *4*(3), 195–226. Accessed Jan. 19, 2010 at http://www.iejme.com/032009/main.htm.

Abrahamson, D. (2010). A tempest in a teapot is but a drop in the ocean: action-objects in analogical mathematical reasoning. In K. Gomez, L. Lyons, & J. Radinsky (Eds.), *Learning in the disciplines: proceedings of the 9th international conference of the learning sciences (ICLS 2010)* (Vol. 1, full papers, pp. 492–499). Chicago: International Society of the Learning Sciences.

Abrahamson, D. (2011). Towards instructional design for grounded mathematics learning: the case of the binomial. In N. L. Stein & S. Raudenbush (Eds.), *Developmental cognitive science goes to school* (pp. 267–281). New York: Taylor & Francis–Routledge.

Abrahamson, D. (2012a). Discovery reconceived: product before process. *For the Learning of Mathematics*, *32*(1), 8–15.

Abrahamson, D. (2012b). Rethinking intensive quantities via guided mediated abduction. *The Journal of the Learning Sciences*, *21*(4), 626–649. doi:10.1080/10508406.2011.633838.

Abrahamson, D., & Cendak, R. M. (2006). The odds of understanding the law of large numbers: a design for grounding intuitive probability in combinatorial analysis. In J. Novotná, H. Moraová, M. Krátká, & N. Stehlíková (Eds.), *Proceedings of the thirtieth conference of the international group for the psychology of mathematics education* (Vol. 2, pp. 1–8). Prague: PME, Charles University.

Abrahamson, D., & Wilensky, U. (2002). *ProbLab*. Evanston: Northwestern University, The Center for Connected Learning and Computer-Based Modeling. http://ccl.northwestern.edu/curriculum/ProbLab/.

Abrahamson, D., & Wilensky, U. (2004). ProbLab: a computer-supported unit in probability and statistics. In M. J. Høines & A. B. Fuglestad (Eds.), *Proceedings of the 28th annual meeting of the international group for the psychology of mathematics education* (Vol. 1, p. 369). Bergen: PME.

Abrahamson, D., & Wilensky, U. (2005a). ProbLab goes to school: design, teaching, and learning of probability with multi-agent interactive computer models. In M. Bosch (Ed.), *Proceedings of the fourth congress of the European Society for Research in Mathematics Education* (pp. 570–579). Barcelona: Universitat Ramon Llull, FUNDEMI IQS.

Abrahamson, D., & Wilensky, U. (2005b). Understanding chance: from student voice to learning supports in a design experiment in the domain of probability. In G. M. Lloyd, M. Wilson, J. L. M. Wilkins, & S. L. Behm (Eds.), *Proceedings of the twenty-seventh annual meeting of the North American chapter of the international group for the psychology of mathematics education* (Vol. 7, pp. 1–8). Roanoke: PME-NA.

Abrahamson, D., & Wilensky, U. (2007). Learning axes and bridging tools in a technology-based design for statistics. *International Journal of Computers for Mathematical Learning*, *12*(1), 23–55.

Abrahamson, D., Trninic, D., Gutiérrez, J. F., Huth, J., & Lee, R. G. (2011). Hooks and shifts: a dialectical study of mediated discovery. *Technology, Knowledge, and Learning*, *16*(1), 55–85.

Abrahamson, D., Gutiérrez, J. F., & Baddorf, A. K. (2012). Try to see it my way: the discursive function of idiosyncratic mathematical metaphor. *Mathematical Thinking and Learning*, *14*(1), 55–80.

Amit, M., & Jan, I. (2007). *A learning environment that promotes concept construction: the case of probability*. Paper presented at the annual conference of the American Educational Research Association, Chicago.

Arnheim, R. (1969). *Visual thinking*. Berkeley: University of California Press.

Artigue, M. (2002). Learning mathematics in a CAS environment: the genesis of a reflection about instrumentation and the dialectics between technical and conceptual work. *International Journal of Computers for Mathematical Learning*, *7*(3), 245–274.

Artigue, M., Cerulli, M., Haspekian, M., & Maracci, M. (2009). Connecting and integrating theoretical frames: the TELMA contribution. In M. Artigue (Ed.), Connecting approaches to technology enhanced learning in mathematics: the TELMA experience [Special issue]. *International Journal of Computers for Mathematical Learning*, *14*, 217–240.

Bamberger, J., & diSessa, A. A. (2003). Music as embodied mathematics: a study of a mutually informing affinity. *International Journal of Computers for Mathematical Learning*, *8*(2), 123–160.

Bamberger, J., & Schön, D. A. (1983). Learning as reflective conversation with materials: notes from work in progress. *Art Education*, *36*(2), 68–73.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577–660.

Bartolini Bussi, M. G., & Mariotti, M. A. (2008). Semiotic mediation in the mathematics classroom: artefacts and signs after a Vygotskian perspective. In L. D. English, M. G. Bartolini Bussi, G. A. Jones, R. Lesh, & D. Tirosh (Eds.), *Handbook of international research in mathematics education* (2nd revised ed., pp. 720–749). Mahwah: Erlbaum.

Barwise, J., & Etchemendy, J. (1991). Visual information and valid reasoning. In W. Zimmerman & S. Cunningham (Eds.), *Visualization in teaching and learning mathematics* (pp. 3–25). USA: Math. Assoc. of America.

Batanero, C., Navarro-Pelayo, V., & Godino, J. D. (1997). Effect of implicity combinatorial model on combinatorial reasoning in secondary school pupils. *Educational Studies in Mathematics*, *32*, 181–199.

Bateson, G. (1972). *Steps to an ecology of mind: a revolutionary approach to man's understanding of himself*. New York: Ballantine Books.

Bereiter, C. (1985). Towards the solution of the learning paradox. *Review of Educational Research*, *55*(2), 210–226.

Borovcnik, M., & Bentz, H.-J. (1991). Empirical research in understanding probability. In R. Kapadia & M. Borovcnik (Eds.), *Chance encounters: probability in education* (pp. 73–105). Dordrecht: Kluwer Academic.

Bruner, J. S. (1960). *The process of education: a searching discussion of school education opening new paths to learning and teaching*. New York: Vintage.

Cheng, P. C.-H. (2011). Probably good diagrams for learning: representational epistemic recodification of probability theory. *Topics in Cognitive Science*, *3*, 475–498.

Cobb, P. & Bauersfeld, H. (Eds.) (1995). *The emergence of mathematical meaning: interaction in classroom cultures*. Hillsdale: Erlbaum.

Cole, M., & Wertsch, J. V. (1996). Beyond the individual-social antinomy in discussions of Piaget and Vygotsky. *Human Development*, *39*(5), 250–256.

Collins, A. (1992). Towards a design science of education. In E. Scanlon & T. O'Shea (Eds.), *New directions in educational technology* (pp. 15–22). Berlin: Springer.

Confrey, J. (2005). The evolution of design studies as methodology. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 135–151). Cambridge: Cambridge University Press.

Davis, P. J. (1993). Visual theorems. *Educational Studies in Mathematics*, *24*, 333–344.

diSessa, A. A. (2008). A note from the editor. *Cognition and Instruction*, *26*(4), 427–429.
Fischbein, E. (1987). *Intuition in science and mathematics*. Dordrecht: Reidel.
Fischbein, E., & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for Research in Mathematics Education*, *28*(1), 96–105.
Freudenthal, H. (1971). Geometry between the devil and the deep sea. *Educational Studies in Mathematics*, *3*(3/4), 413–435.
Freudenthal, H. (1974). The crux of course design in probability. *Educational Studies in Mathematics*, *5*(3), 261–277.
Fuson, K. C. (2009). Avoiding misinterpretations of Piaget and Vygotsky: mathematical teaching without learning, learning without teaching, or helpful learning-path teaching? *Cognitive Development*, *24*(4), 343–361. doi:10.1016/j.cogdev.2009.09.009.
Gelman, R. (1998). Domain specificity in cognitive development: universals and nonuniversals. In M. Sabourin, F. Craik, & M. Robert (Eds.), *Advances in psychological science: Vol. 2. Biological and cognitive aspects* (pp. 50–63). Hove: Psychology Press.
Getzels, J. W., & Csikszentmihalyi, M. (1976). *The creative vision: a longitudinal study of problem finding in art*. New York: Wiley.
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.
Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: why biased minds make better inferences. *Topics in Cognitive Science*, *1*(1), 107–144.
Glenberg, A. M., & Robertson, D. A. (1999). Indexical understanding of instructions. *Discourse Processes*, *28*, 1–26.
Goldin, G. A. (1987). Levels of language in mathematical problem solving. In C. Janvier (Ed.), *Problems of representation in the teaching and learning of mathematics* (pp. 59–65). Hillsdale: Erlbaum.
Hadamard, J. (1945). *The psychology of invention in the mathematical field*. New York: Dover.
Harel, G. (2013). Intellectual need in mathematical practice: a theoretical perspective. In K. Leatham (Ed.), *Vital directions for mathematics education research* (pp. 119–152). New York: Springer.
Hoffmann, M. H. G. (2003). Peirce's 'diagrammatic reasoning' as a solution of the learning paradox. In G. Debrock (Ed.), *Process pragmatism: essays on a quiet philosophical revolution* (pp. 121–143). Amsterdam: Rodopi.
Iversen, K., & Nilsson, P. (2007). Students' reasoning about one-object stochastic phenomena in an ICT-environment. *International Journal of Computers for Mathematical Learning*, *12*(2), 113–133.
Johnson, M. L. (1999). Embodied reason. In G. Weiss & H. F. Haber (Eds.), *Perspectives on embodiment* (pp. 81–102). New York: Routledge.
Jones, G. A., Langrall, C. W., & Mooney, E. S. (2007). Research in probability: responding to classroom realities. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 909–955). Charlotte: Information Age Publishing.
Kahneman, D., Slovic, P., & Tversky, A. (Eds.) (1982). *Judgment under uncertainty: heuristics and biases*. Cambridge: Harvard University Press.
Kamii, C. K., & DeClark, G. (1985). *Young children reinvent arithmetic: implications of Piaget's theory*. New York: Teachers College Press.
Karmiloff-Smith, A. (1988). The child is a theoretician, not an inductivist. *Mind & Language*, *3*(3), 183–195.
Kirsh, D. (1996). Adapting the environment instead of oneself. *Adaptive Behavior*, *4*(3–4), 415–452.
Konold, C. (2004). TinkerPlots: dynamic data exploration: overview—empowering students as data analysts from http://www.umass.edu/srri/serg/projects/tp/tpovervw.html.
Koschmann, T., Kuuti, K., & Hickman, L. (1998). The concept of breakdown in Heidegger, Leont'ev, and Dewey and its implications for education. *Mind, Culture and Activity*, *5*(1), 25–41.
Lakoff, G., & Johnson, M. L. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Lakoff, G., & Núñez, R. E. (2000). *Where mathematics comes from: how the embodied mind brings mathematics into being*. New York: Basic Books.

Mariotti, M. A. (2009). Artifacts and signs after a Vygotskian perspective: the role of the teacher. *ZDM: The International Journal on Mathematics Education*, *41*, 427–440.

Newman, D., Griffin, P., & Cole, M. (1989). *The construction zone: working for cognitive change in school*. New York: Cambridge University Press.

Núñez, R. E., Edwards, L. D., & Matos, J. F. (1999). Embodied cognition as grounding for situatedness and context in mathematics education. *Educational Studies in Mathematics*, *39*, 45–65.

Pascual-Leone, J. (1996). Vygotsky, Piaget, and the problems of Plato. *Swiss Journal of Psychology*, *55*(2/3), 84–92.

Peirce, C. S. (1867). Upon the logic of mathematics. *Proceedings of the American Academy of Arts and Sciences*, *7*, 402–412.

Piaget, J. (1968). *Genetic epistemology* (trans: Duckworth, E.). New York: Columbia University Press.

Piaget, J., & Inhelder, B. (1969). *The psychology of the child*. New York: Basic Books. (Trans: Weaver, H. Original work published 1966).

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

Prediger, S. (2008). Do you want me to do it with probability or with my normal thinking? Horizontal and vertical views on the formation of stochastic conceptions. *International Electronic Journal of Mathematics Education*, *3*(3), 126–154.

Pylyshyn, Z. (1973). What the mind's eye tells the mind's brain: a critique of mental imagery. *Psychological Bulletin*, *80*, 1–24.

Radford, L. (2003). Gestures, speech, and the sprouting of signs: a semiotic–cultural approach to students' types of generalization. *Mathematical Thinking and Learning*, *5*(1), 37–70.

Rubel, L. (2009). Student thinking about effects of sample size: an in and out of school perspective. In S. L. Swars, D. W. Stinson, & S. Lemons-Smith (Eds.), *Proceedings of the thirty-first annual meeting of the North American chapter of the International Group for the Psychology of Mathematics Education* (Vol. 5, pp. 636–643). Atlanta: Georgia State University.

Saxe, G. B. (2004). Practices of quantification from a sociocultural perspective. In K. A. Demetriou & A. Raftopoulos (Eds.), *Developmental change: theories, models, and measurement* (pp. 241–263). New York: Cambridge University Press.

Schön, D. A. (1983). *The reflective practitioner: how professionals think in action*. New York: Basic Books.

Sfard, A. (2002). The interplay of intimations and implementations: generating new discourse with new symbolic tools. *The Journal of the Learning Sciences*, *11*(2&3), 319–357.

Sfard, A. (2007). When the rules of discourse change, but nobody tells you—making sense of mathematics learning from a commognitive standpoint. *The Journal of the Learning Sciences*, *16*(4), 567–615.

Sfard, A. & McClain, K. (Eds.) (2002). Analyzing tools: perspectives on the role of designed artifacts in mathematics learning [Special Issue]. *Journal of the Learning Sciences*, *11*(2&3).

Shaughnessy, J. M. (1977). Misconceptions of probability: an experiment with a small-group, activity-based model building approach to introductory probability. *Educational Studies in Mathematics*, *8*, 295–316.

Smith, J. P., diSessa, A. A., & Roschelle, J. (1993). Misconceptions reconceived: a constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences*, *3*(2), 115–163.

Steinbring, H. (1991). The theoretical nature of probability in the classroom. In R. Kapadia & M. Borovcnik (Eds.), *Chance encounters: probability in education* (pp. 135–168). Dordrecht: Kluwer Academic.

Stetsenko, A. (2002). Commentary: Sociocultural activity as a unit of analysis: how Vygotsky and Piaget converge in empirical research on collaborative cognition. In D. J. Bearison & B. Dorval (Eds.), *Collaborative cognition: children negotiating ways of knowing* (pp. 123–135). Westport: Ablex Publishing.

Stohl, H., & Tarr, J. E. (2002). Developing notions of inference with probability simulation tools. *The Journal of Mathematical Behavior*, *21*(3), 319–337.

Suzuki, S., & Cavanagh, P. (1998). A shape-contrast effect for briefly presented stimuli. *Journal of Experimental Psychology. Human Perception and Performance*, *24*(5), 1315–1341.

Vagle, M. D. (2010). Re-framing Schön's call for a phenomenology of practice: a post-intentional approach. *Reflective Practice*, *11*(3), 393–407. doi:10.1080/14623943.2010.487375.

Vérillon, P., & Rabardel, P. (1995). Cognition and artifacts: a contribution to the study of thought in relation to instrumented activity. *European Journal of Psychology of Education*, *10*(1), 77–101.

von Glasersfeld, E. (1987). Learning as a constructive activity. In C. Janvier (Ed.), *Problems of representation in the teaching and learning of mathematics* (pp. 3–18). Hillsdale: Erlbaum.

Vos Savant, M. (1996). *The power of logical thinking: easy lessons in the art of reasoning... and hard facts about its absence in our lives*. New York: St. Martin's Griffin.

Wertsch, J. V. (1985). *Vygotsky and the social formation of mind*. Cambridge: Harvard University Press.

Wilensky, U. (1993). *Connected mathematics—building concrete relationships with mathematical knowledge*. Cambridge: MIT. http://ccl.northwestern.edu/papers/download/Wilensky-thesis.pdf.

Wilensky, U. (1995). Paradox, programming and learning probability. *The Journal of Mathematical Behavior*, *14*(2), 231–280.

Wilensky, U. (1996). Modeling rugby: kick first, generalize later? *International Journal of Computers for Mathematical Learning*, *1*(1), 125–131.

Wilensky, U. (1997). What is normal anyway? Therapy for epistemological anxiety. *Educational Studies in Mathematics*, *33*(2), 171–202.

Wilensky, U., & Papert, S. (2010). Restructurations: reformulations of knowledge disciplines through new representational forms. In J. Clayson & I. Kallas (Eds.), *Proceedings of the constructionism 2010 conference*, Paris.

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(13), 5012–5015.

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

Sticking to Your Guns: A Flawed Heuristic for Probabilistic Decision-Making

Deborah Bennett

Abstract The much-investigated Monty Hall problem is often regarded as a testament to the woeful state of probabilistic thinking. But there is some evidence that individuals continue to refuse to switch doors with Monty (even in light of additional evidence indicating that their probability of winning has doubled) for reasons that are not related to probabilistic thinking at all. The steadfast refusal to switch can be described as a gut reaction rather than a rational reaction. Having the semblance of superstitions, intuitions involving sticking with your gut are common—refusing to change lanes from a frustratingly slow lane to another lane that is moving briskly (car lanes or grocery lines), refusing to deviate from your first answer choice in a multiple-choice test even upon thoughtful reconsideration, or refusing to exchange lottery tickets even when offered a reward to switch. This study examines the development of a heuristic or rule of thumb about (not) switching with regards to cognitive biases as put forward by decision theorists. These biases influence our thinking, cause us to make incorrect assumptions, and often result in our employing suboptimal strategies that may seem irrational. In this research, experimental treatments allow us to examine how individuals make simple probabilistic decisions when the element of sticking with your first choice is inserted or removed. Additionally, I attempt to examine whether individuals are of "two minds," whether some decisions are based on gut instinct and other decisions are based on rationality. In light of these findings, mathematics educators may be better prepared take cognitive biases into consideration when assessing their students' probabilistic thinking and decisions.

The Monty Hall Problem asks you to imagine that you are a contestant on the television show, *Let's Make a Deal*. Its host, Monty Hall, asks you to choose one of three doors in order to receive the "prize" behind it. Behind two doors are goats and behind a third door is a car; Monty knows what is where and never moves "prizes" to another door once you have chosen your door. After you have chosen a door, Monty

D. Bennett (✉)
New Jersey City University, 2039 Kennedy Blvd., Jersey City, NJ, USA
e-mail: dbennett@njcu.edu

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_14,

opens one of the other two doors to reveal a goat. Now Monty wants to make a deal. Will you swap your door for Monty's remaining unopened door?

The optimal strategy is to take Monty's deal and switch from your first choice to Monty's unopened door—switching wins two-thirds of the time. Most individuals think that the chances are 50–50 that the car is behind the door originally selected by the contestant and Monty's remaining unopened door. Part of the difficulty individuals have with this problem lies in failing to recognize that the probabilities are no longer equally likely. Originally, each of the three doors was equally likely to have the car behind it, but once Monty opened one of his doors with a goat behind it, the remaining unopened doors—yours and his—are not equally likely to have a car behind them. In fact, Monty's door is twice as likely to have the car behind it. Your original chances of choosing the door with the car were $\frac{1}{3}$ versus a $\frac{2}{3}$ chance that the car was behind one of the other two doors. The chance that the car is behind your door is still $\frac{1}{3}$, leaving a $\frac{2}{3}$ chance that the car is behind Monty's remaining unopened door instead of yours. Another part of the difficulty individuals have with this problem lies in not realizing the impact of the fact that Monty's choice of which door to open is not haphazard.[1] Monty knows where the car is and *always* opens a door with a goat behind it. If Monty did make a random choice of which of his two doors to open, once a door was opened (and if a goat was behind it) then the chances of which unopened door had the car behind it would be 50–50.

Prior research on the Monty Hall Problem indicates that people rarely solved the problem correctly; they allow their intuition to trump rules of logic (Franco-Watkins et al. 2003). After Monty has shown a goat behind one door, people think that the odds of their own door or the other remaining door are the same (wrong), but they overwhelmingly stick with their original door. They are committed to *their* door, and this propensity to stay with one's original door has been demonstrated in a variety of cultures across the world.

In study after study, even though individuals believe (wrongly) that the odds are even and therefore it doesn't matter which door the contestant chooses, contestants overwhelmingly refuse to swap doors with Monty. This, then, seems to have very little to do with actual probabilistic thinking, and more to do with a kind of gut reaction. My conjecture was that this entrenched superstition of going with your initial choice interferes with making a rational decision in probability problems such as these.

I theorized that the refusal to give up one's original door has to do with feeling as if you now own the door—it is your door. Behavioral economists call this the *endowment effect* and have shown that individuals attribute greater value to an object once they own it. Furthermore, economists and decision theorists remind us that individuals are *loss averse*. People prefer avoiding losses to acquiring gains. And then there is *anticipatory regret*. We seek to avoid regret and we anticipate much

[1]Krauss and Wang (2003) have referred to this as the *ambiguity of Monty's strategy* (p. 12). However, their review of the history of the problem and the research of others seems to indicate that individuals have difficulty with the problem of switching even when Monty's strategy is made explicit.

regret if we abandon our first choice, and it is, after all, the winning door. (Even the word "abandon" is negatively charged.)

Should we switch and find our initial choice to be a winner, we anticipate "kicking ourselves." If only we hadn't switched doors—a kind of *counterfactual thinking* according to psychologists. Refusal to switch doors, even in the light of increased odds of winning, is a type of *status quo bias*. Individuals are hesitant to deviate from the status quo. Linked to this bias is another bias, the preference for inaction over action. Individuals hold themselves more responsible for negative outcomes due to their actions than similar or identical outcomes due to their inactions. The sins of *commission* (acting) seem to be greater than the sins of *omission* (not acting). All of these cognitive biases can lead to a sort of *memory distortion*, and researchers have documented that counterfactual thoughts and anticipatory regret affect people's decision strategies. Our mental sample space is distorted by an overrepresentation of negative results attached to losses and regret. Their vividness makes them seem more frequent than they really are. There is some evidence that believing in good luck imbues individuals with a sense of control and confidence. If so, then perhaps in tempting fate by switching doors we feel a loss of control; we have invited disaster and angered the gods. Finally, Risen and Gilovich (2007, 2008) conclude that individuals are of "two minds" in making everyday judgments about likelihoods—what their guts tell them and what their rational minds tell them.

I tested my conjecture by giving participants a simplified version of the Monty Hall Problem: after making an initial choice, new information was revealed that didn't require an omniscient host. The first and second treatments involved two different scenarios where individuals made an initial choice, received additional information, and then were asked whether they would choose to stick with their original choice or switch. In other words, do they go with their gut or not? The first treatment was one in which the participant's first choice did not have the higher probability of winning and the second treatment was one in which the participant's first choice did have the higher probability of winning. So in the first treatment, sticking was not the optimal strategy, and in the second, sticking was the optimal strategy. The third and forth treatments were scenarios identical to the first and second treatments, respectively, but the final question implied that there was a rational decision to be made and the responses allowed for three rational choices: sticking with your original choice, switching, or deciding that both choices give equally likely results. In a fifth treatment, the initial choice was removed from the scenario altogether, eliminating the temptation to stick with the first choice: the scene was set, the new information revealed, and only then were the participants allowed to make a choice.

1 Background

1.1 Endowment Effect

Research on why individuals refuse to switch doors with Monty even though they believe that the doors are equally likely to hide a car has a parallel in research on why individuals refuse to exchange lottery tickets for another ticket equally likely to be a winner (even when an incentive is offered). Risen and Gilovich have studied the phenomenon of people refusing to exchange lottery tickets and concluded that people do this for the same reasons that students are reluctant to change their initial answers on a multiple-choice test. Furthermore, they are the same reasons that individuals refuse to switch from a frustratingly slow grocery line to a faster one, the same reasons that individuals are reluctant to switch lanes in traffic to one that is moving,[2] and the same reasons that individuals will not exchange doors in the Monty Hall problem (Risen and Gilovich 2007, 2008).

Early studies by Knetsch and Sinden of the endowment effect demonstrated that owners are reluctant to part with their endowment (as cited by Kahneman et al. 1991). In particular, participants, once endowed with a lottery ticket, were reluctant to exchange them for sure money. After more than a decade of research on this topic, Kahneman, Knetsch, and Thaler concluded that the endowment effect is not due to the appeal of the object owned (the lottery ticket) but rather, the pain of giving it up. "Forgone gains are less painful than perceived losses" (Kahneman et al. 1991, p. 203). This pain of loss has also been linked to the tendency to remain at the status quo, saying that "the disadvantages of leaving [the status quo] loom larger than advantages" (Kahneman et al. 1991, pp. 197–198).

This asymmetry of feelings for gains and losses was termed *loss aversion* by Kahneman and Tversky in 1984 (as cited by Kahneman et al. 1991). The reluctance to exchange a lottery ticket is a type of status quo bias and some status quo biases are attributed to loss aversion. However, Risen and Gilovich (2007) claimed that unlike other status quo biases, this reluctance cannot be attributed to loss aversion, but rather, to *regret avoidance*. Individuals imagine they may wind up blaming themselves for mistaken actions, and they anticipate these feelings.

1.2 "If Only" Counterfactual Thinking

This imagining of blame and self-recrimination is sometimes described as counterfactual thinking, "if only I hadn't exchanged my lottery ticket" and "if only I hadn't switched doors with Monty" (Kruger et al. 2005; Miller and Taylor 1995; Petrocelli and Harris 2011). Kruger et al. (2005) investigated what they termed, the *First Instinct Fallacy* (FIF), and the counterfactual *if only* thinking that is anticipated in this mistaken belief. The widely held belief that one should not change the answer to a test item because one's first impression is intuitively more accurate is myth and has been debunked by a number of studies since the 1920s (Benjamin et al. 1984; Edwards and Marshall 1977; Skinner 1983). Benjamin, Cavell, and Shallenberger's metasynthesis (1984) reviewed 33 studies from 1928 to 1983—studies that had researched answer-changing beliefs and behaviors on objective tests. Virtually all of

[2]However, it is easier to change your mind and "butt back in" to your original lane when driving. That would be frowned upon in the grocery line.

the studies indicated that the academic folk wisdom that one's first instinct is best is a fallacy. The research studies agreed in the following respects: (i) only a small percentage of test items were ever changed, (ii) most changes reflected a correction from a wrong answer to a right one, and (iii) most changers improved test scores. In terms of attitudes, the studies consistently found that the majority of test-takers believed that answer changing would not improve their scores. In fact, many believed that it would lower their scores (Benjamin et al. 1984).

In an effort to determine where this widespread misconception comes from, Benjamin et al. (1984) surveyed University faculty and found that most faculty were unaware of the studies in this area. In their survey, most faculty (55.2 %) believed that changing the initial answer on an objective test would lower scores; only 15.5 % responded that answer changing would improve a student's score. Even worse, some of the faculty (20.6 %) went so far as to warn their students not to change their initial answers because that would likely lower their grade.

The vast majority of research indicated that the majority of answer changes are from incorrect to correct, but data indicated that 3 out of 4 college students believed that answer-changing lowers scores (Kruger et al. 2005, p. 725). The myth must indeed be difficult to debunk. As recently as 2000, *Barron's* Graduate Records Examination (GRE) Preparation booklet provided a warning against answer-changing as advice to potential test-takers (as cited by Kruger et al. 2005).

Even when subjects were provided historical data on the advantages of answer-changing, behavior was not noticeably altered on subsequent tests (see recent studies cited in Kruger et al. 2005). The difficulty in debunking belief in a myth like this may lie in the selective memory of the test-takers. Bath hypothesized that students "seldom remember those items that they change and got correct" (as cited in Benjamin et al. 1984), but they vividly recall the times they changed to a wrong answer and their first answer was right (they were kicking themselves).

Writing about counterfactual thought, regret, and superstition, Miller and Taylor observed that the belief that one should stick with one's first instinct on a multiple-choice test is "widely held by students and strongly endorsed by professors and test consultants" (Miller and Taylor 1995, p. 309). They proposed the following:

> If the decision to change a correct answer to an incorrect answer (an error of commission) leads to more if only self-recriminations than does the decision not to change an incorrect answer (an error of omission), instances of the former can be predicted to be more available in students' memories. This differential availability will provide them with compelling personal evidence for the wisdom of following their first instinct (Miller and Taylor 1995, p. 309).

This anticipation of self-recrimination has been reported in many studies involving "Monty Hall"-type decisions (Kruger et al. 2005; Petrocelli and Harris 2011). Folks tend not to change from the status quo unless there is a compelling reason, and the disadvantages of leaving the status quo threaten and menace. Kahneman et al. (1991) proposed, "Changes that make things worse (losses) loom larger than improvements or gains" (p. 199). But it is not simply the change from the status quo that is anticipated as painful, it is being responsible for *initiating* the change that is particularly painful.

1.3 Preference for Inaction

Individuals tend to favor the status quo over equally attractive alternatives and they prefer inaction to action (Ritov and Baron 1992). In other words, they prefer omission (not acting) to commission (acting). They also react more strongly to adverse outcomes resulting from change than those resulting from the status quo. Ritov and Baron (1992) attempted to separate the effects of preferences for maintaining the current state from preferences for inaction. Their results indicated that the primary bias is toward inaction (omission) rather than toward keeping things the same. Their subjects preferred the inaction alternative, independent of whether inaction would lead to keeping the status quo or not. "Negative consequences of the commission are weighed more heavily than positive consequences of the commission, regardless of the status quo" (Ritov and Baron 1992, p. 11).

Miller and Taylor (1995) proposed:

> Negative events preceded by acts of commission (e.g., being delayed after switching lines) are more likely to give rise to counterfactual, if only self-recriminations than either negative events preceded by acts of omission (e.g., being delayed after contemplating but deciding against switching lines) or positive outcomes preceded by acts of either commission or omission (e.g., being accelerated after either switching or not switching lines) (Miller and Taylor 1995, p. 307).

Others have found that it is easier to generate a counterfactual "what if" thought based on the deletion of an action that did take place (if only I hadn't done that) than it is to imagine an action that didn't take place (if only I had done any number of other things) (Roese and Olson 1995).

In his blackjack experiments, Taylor found that subjects recalled losses following acts of commission more often than losses following acts of omission even when they were less numerous (cited in Miller and Taylor 1995, p. 313). Not only are negative experiences more available in memory, avoidable negative experiences are more likely to be remembered than unavoidable ones. They stand out in our memory.

More generally, studies have found "events preceded by actions are more easily imagined otherwise" than events preceded by inaction (Kruger et al. 2005, p. 726; Miller and Taylor 1995). An error that results from an action (for example, changing a correct answer to an incorrect one) seems like an error that "almost did not happen" and therefore could have been easily avoided. Kruger, Wirtz, and Miller hypothesized that in considering whether to change lines in a grocery store (or lanes on the highway) we easily envision the line we change to slowing down and our original line speeding up. They argued that changing lines when one should not have is more frustrating and memorable than failing to change lines when one should have, so the occurrences of these events are more vivid in memories. We suffer more regret when our own actions cause our misfortunes.

Regret and self-blame arise "whenever a person has precipitated a negative outcome by doing something he or she can easily imagine not having done" (Miller and Taylor 1995, p. 322)—when one deviates from one's customary practice. "The asymmetry [between omission and commission] affects both blame and regret after a mishap, and the *anticipation* [emphasis mine] of blame and regret, in turn, could affect behavior" (Kahneman et al. 1991, p. 202).

1.4 First Instinct Fallacy and Tempting Fate

Research on the first instinct fallacy found that people overestimate the effectiveness of sticking with their first answer and that switching from a correct answer to an incorrect one is more irksome that failing to switch to a correct answer from an incorrect one. Furthermore, there is an asymmetry in the way we remember the effectiveness of the strategy to not switch or switch (not switching is remembered as a better strategy than it actually is and switching is remembered as a worse strategy than it really is). Kruger et al. (2005) concluded that the asymmetry in frustration causes an asymmetry in memory (link between regret and memory) which in turn confirms and enhances the first instinct fallacy. They concluded that another factor may be "fear of feeling foolish" (Kruger et al. 2005, p. 733). We feel more foolish taking action when we shouldn't have than sticking with the status quo and accepting our fate. Risen and Gilovich (2008) concluded that in addition to the negative outcomes that are prominent in our memory there is an additional negative dose of humiliation and regret. People are tormented by counterfactual thought: I wish I hadn't done that; I wish I hadn't said that. Our actions are something we can control.

Anyone who drives a regular traffic route when there are alternate routes available has had the experience: You take a route different from your usual one and get into an enormous traffic jam. In hindsight you say to yourself, "I knew I shouldn't have come by this route. I should not have changed my pattern." We feel greater responsibility for undesirable outcomes when they result from changes we initiated. With no action on our part, it cannot be our fault—it is fate.

Risen and Gilovich have characterized our reluctance to switch from our first answer on a multiple-choice test, our refusal to switch to a faster moving grocery line, and our reluctance to exchange lottery tickets as a reluctance to tempt fate. Regret is particularly keen when one tempts fate and a negative outcome occurs because we may well feel that by disobeying a shared rule we "deserve it." "Once a given superstition gains some acceptance in a social group, no matter how arbitrary (don't walk under a ladder, don't comment on success), the thought of flaunting it makes the prospect of a negative outcome seem especially negative and ... especially likely" (Risen and Gilovich 2008, p. 303).

Risen and Gilovich (2007) contended that the reluctance to tempt fate plays a prominent role in the reluctance to exchange lottery tickets. If you anger the gods, bad things will happen to you. This creates a conflict between what one rationally knows (in the case of lottery tickets, the odds are the same and won't change) and an intuitive sense that the exchanged ticket would be more likely to win. In fact, Risen and Gilovich (2007) found that subjects believed that the *very act of exchanging* the lottery ticket made it more likely to win, and this belief was persistent.

Actions that tempt fate are sometimes considered acts of greed (selfishness), needless risk (rash action), and hubris. Counterfactuals and feelings of regret most often arise when the negative outcome "fits the crime." Risen and Gilovich contended that we imagine or anticipate negative outcomes that punish the fate-tempting behavior. We think that the universe is "interested not only in punishing certain behaviors but in punishing them a certain, ironic way" (Risen and Gilovich 2008, p. 304). *What goes around, comes around.*

1.5 Memory and Regret

Miller and Taylor (1995) proposed two routes by which counterfactual thoughts "generated by a highly mutable event sequence" can lead to superstitious beliefs and affect people's decision strategies: *memory distortion* (the greater number of "if only" thoughts after ill-fated events serves to make these thoughts more available in memory "and hence more subjectively probable" p. 306), and *anticipatory regret*—people's "reluctance to engage in actions easily imagined otherwise." Memory distortion links two processes described by Tversky and Kahneman in their 1973 *availability* heuristic: the process by which events become available in memory and the process by which events become available in imagination (Tversky and Kahneman 1973). Miller and Taylor (1995) also stated that "event sequences that yield highly available alternative constructions" tend to be highly available in memory. Anticipatory regret predicts that any bad thing that does happen by "tempting fate" or flaunting superstition (ill-fated acts) will be psychologically more painful. Additionally, the greater self-recriminations following unfortunate acts of commission make those experiences more prominent in our memories and their greater prominence in memory causes people to overestimate their commonness.

Miller and Taylor attempted to separate the roles played by memory distortion and anticipatory regret in the development of superstitious belief, but they admitted that this is difficult since they often operate in tandem. They explained common intuitions that have the "feel" of superstitions—like the idea that switching lines (to a faster line) in the grocery store "virtually ensures that one's old line will speed up and one's new line will slow down" (Miller and Taylor 1995, p. 307).

Petrocelli and Harris (2011) replicated the Monty Hall Problem with subjects as they attempted to learn why people experience such difficulty learning the switch concept rule even after multiple trials of the problem. They found that counterfactual thinking was most pronounced in relation to losses that occurred after switching. A memory bias against switching appeared to inhibit learning and increase the likelihood of adhering to a long-term losing strategy that seems irrational. The counterfactual thoughts (kicking oneself) that emerged in response to switch losses were due to the fact that it is easier to imagine oneself winning if only one had not switched. We have stronger regrets for our actions over our inactions (if only I hadn't switched doors), and we feel greater responsibility for undesirable outcomes that result from decisions we make to change things. In the Monty Hall Problem, the "sting" of losses from switching was greater than that of losses from sticking, and being more vivid, the instances were overrepresented in memory.

In addition to anticipated regret, Risen and Gilovich (2007) offered another reason for this reluctance to switch doors in the Monty Hall Problem: Individuals may believe that the odds of a positive result are better if they stick than if they switch. People's memories for the frequencies of different sequences affect their beliefs about the likely consequences of different decision strategies (Miller and Taylor 1995, p. 321).

According to Miller and Taylor, the bias of our subjective expectancies is only one of the ways that counterfactual regret can affect decision-making strategies (Miller and Taylor 1995, p. 314). "If only" regret (kicking yourself) can impact the decision maker's utility function—that metric we use to determine how different possible outcomes are valuated. It is not necessary to conclude that individuals have a biased memory; they may be correctly anticipating the psychological pain that a particular action will cause. Miller and Taylor reason, "If a student anticipates being tormented by having changed a correct answer to a wrong one, he or she might reasonably ask: Why risk it?" (Miller and Taylor 1995, p. 315). Some individuals might just prefer to lose through inaction rather than through action (Miller and Taylor 1995, p. 321). By deciding not to act, we accept our fate. By acting, we may tempt fate.

Miller and Taylor (1995) stated, "One is just 'asking for it' if one takes an action he or she knows, in retrospect, will be easy to imagine otherwise" (p. 326). They cited the example of wearing your lucky shoes to your final exam. Just because you wear them doesn't mean that you really believe they will increase your chances of doing better, nor does it mean that you believe that in not wearing your lucky shoes you will do worse, you may just feel, why tempt fate?

1.6 System 1 and System 2 Processes

Just as you can be of two minds about the effectiveness of wearing your lucky shoes, many propose that we are of two minds in making judgments involving chance (Franco-Watkins et al. 2003; Kahneman 2003; Petrocelli and Harris 2011; Risen and Gilovich 2007, 2008). Risen and Gilovich propose that even though our rational minds tell us that there is no way that a sports announcer mentioning a no-hitter in progress can change the probability of the game ending in a no-hitter, our gut tells us not to tempt fate. Our reluctance to tempt fate leads us to remain in an exasperatingly slow checkout line at the grocery store. We may know in our rational minds that believing that the moment we leave our line to join the speedier line, the new line will slow down and the original line will speed up is irrational, but we stay in our original line nonetheless (Risen and Gilovich 2008).

Basing his article on his Nobel Prize lecture, Daniel Kahneman wrote in 2003 about two systems of cognition: intuition and reasoning.[3] System 1 processes (intuition) are "thoughts and preferences that come to mind quickly and without much

[3] Kahneman also remarks that the work revisits much of his previous work with Amos Tversky and draws extensively on collaboration with Shane Federick in 2002.

reflection" (Kahneman 2003, p. 697) while System 2 (reasoning) is a deliberate, slower thought process. System 1 operations are born from habit and are difficult to control or modify. System 2 operations are governed by rules. Kahneman notes that one of the functions of System 1 is to provide a natural or intuitive assessment of stimuli as good or bad. (It is bad to tempt fate. It is good to stick with your first answer.)

System 1 intuitive decisions are influenced by the accessibility of features of the situation. System 2 monitors System 1 and can modify or override an intuition, but according to Kahneman, the monitoring is "quite lax and allows many intuitive judgments to be expressed, including some that are erroneous" (p. 699).

Risen and Gilovich considered that the belief that the universe will punish us for tempting fate was best explained by a dual-process or two-systems perspective. "The intuitive system believes [that the universe will punish them] and the rational system does not" (Risen and Gilovich 2008, p. 301). They maintained that belief that negative outcomes were more likely following behavior that tempted fate increased when the System 2 monitoring of System 1 was inhibited.

Individuals are "of two minds" in making everyday judgments about likelihoods. When asked to base their judgment on rationality (probability scales) rather than subjective likelihoods, individuals were more likely to rely on rational thought processes and less inclined to go with their gut. However, Risen and Gilovich (2007) concluded that more often than not it is the intuitive systems that people obey.

Petrocelli and Harris (2011) stated, "The default heuristic used to approach the [Monty Hall] problem competes with more deliberate and logical processing" (p. 12). Furthermore, they represented that the competition between the default heuristic and the rational process generalized to other decisions that involve making a revocable decision, receiving additional information, and ultimately making a final determination. In these decisions too, a clearly optimal strategy may be neglected for a suboptimal one (Petrocelli and Harris 2011, p. 13).

1.7 Summary

The endowment effect appears to affect probabilistic decision-making. Furthermore, our tendency for loss aversion and regret avoidance seem to strengthen the endowment effect. We opt for the status quo, and we see reaction as less painful than action when there are negative consequences. We believe in first instincts and sticking to our guns.

2 Method

The background research suggested several research questions:

- Does the endowment effect instill decision makers with a sense of ownership of their first choice even when their first choice is random?

- If so, is there a difference between sticking and switching behavior when one's first choice is less than optimal versus an optimal first choice?
- Can rational decisions (System 2) modify the gut or intuitive decisions (System 1) of sticking with one's first choice?
- When the endowment effect is removed, is probabilistic thinking improved?

In order to investigate these questions, a study was designed, involving decisions to stick with one's original decision or switch, without the confounding effect of Monty Hall's omniscience.

Participants were selected from among college students at New Jersey City University taking a freshman algebra or freshman psychology course. Both courses were part of the general education curriculum (without college-level prerequisites) and were taken by students in all different majors. It was assumed that the students had no particular mathematics or probability expertise. Participation was voluntary and did not contribute or detract from the students' course grades in any way. The five treatments were distributed in all classes. Each of the 163 participants was given one treatment only and one question to answer. There was an opportunity for participants to volunteer for further involvement in the study should qualitative data on how participants came to make their decisions be gathered later. Participants were guaranteed anonymity if they ended their participation after answering the one question (and confidentiality if they volunteered to be contacted later). Participants were told only that the researcher was interested in how individuals make decisions. All treatments involved the choice between one of two envelopes that had money in them.[4]

In *Treatments 1* and *3* the scenarios were the same; the question asked and the answers allowed were different. In both cases, the participant's decision to stick with his or her original choice was the "wrong" decision.[5] *Treatment 1* was the gut choice version—forcing the participant to make a (gut) choice to stick or switch and *Treatment 3* suggested that there was a rational choice to be made and allowed the participant to choose a third option.

Scenario for *Treatments 1 and 3*

I have two identical envelopes. You watch me put two \$1 bills in one envelope and, in the other, I put a \$1 bill and a \$100 bill. We will create a situation where you may keep one envelope once I give you the choice.

Behind my back, I mix the envelopes and then allow you to choose an envelope. The one you chose, we will call your envelope and the other we'll call mine. I ask you to randomly select a bill from your envelope without looking. We see that the bill is a \$1 bill. You return it to its envelope. You may now choose an envelope to keep, yours or mine.

In this scenario, there is a probability of $\frac{2}{3}$ that the \$100 bill is in my envelope and $\frac{1}{3}$ that the \$100 bill is in the participant's envelope, so the "correct" decision

[4]Although this experiment involved two envelopes with money in them, it is not the famous Two Envelope Paradox or Exchange Paradox that involves infinite expected value.

[5]Hereafter, I refer to the probabilistic choice that optimized the participants' expected returns as the "correct" decision and the less optimal choice as the "wrong" decision.

to maximize the participant's expected return is to exchange his or her envelope for mine. The correct solution can be worked out using conditional probabilities and Bayes' formula, but an informal analysis can be used by asking yourself the following question: Now that I have seen a $1 bill removed at random from my envelope, which envelope is mine more likely to be—the envelope with two $1 bills or the envelope with one $1 bill (and another $100 bill)? The questions for *Treatments 1* and *3* were different.

Treatment 1

What will you decide—to keep yours or to exchange it for mine? (Circle one answer.)

(a) Keep
(b) Exchange

Treatment 3

What should you decide—to keep yours or to exchange it for mine? (Circle one answer.)

(a) Keep
(b) Exchange
(c) It doesn't matter

In both treatments, the decision to keep one's original envelope (stick with one's first choice) was *inconsistent* with the optimal decision. *Treatment 1* (*gut*) forces the participant to make a choice to keep or exchange, while the question in *Treatment 3* (*rational*) hints that there is a responsible decision ("what *should* you decide") and also attempts to capture the responses of participants who actually think the outcomes are 50–50 by offering a third option, that it doesn't matter whether they keep or exchange. Thirty-two (32) students participated in each treatment.

In *Treatment 2* and *Treatment 4*, the decision to stick with one's first choice is *consistent* with the optimal decision.

Scenario for *Treatments 2 and 4*

I have two identical envelopes. You watch me put two $1 bills in one envelope and, in the other, I put a $1 bill and a $100 bill. We will create a situation where you may keep one envelope once I give you the choice.
Behind my back, I mix the envelopes and then allow you to choose an envelope. The one you chose, we will call your envelope and the other we'll call mine. I ask you to randomly select a bill from my envelope without looking. We see that the bill is a $1 bill. You return it to its envelope. You may now choose an envelope to keep, yours or mine.

In this scenario, the decision to stick with your original envelope is the "correct" decision. Other than the fact that the "seen" $1 bill comes from my envelope the scenario is identical to that of *Treatments 1* and *3*. The probability that the $100 bill is in my envelope is $\frac{1}{3}$ and the probability it is in your envelope is $\frac{2}{3}$ so you should not exchange with me; you should keep your envelope. The questions and answers for *Treatment 2* and *Treatment 4* parallel those of *Treatment 1* and *Treatment 3*, respectively.

Treatment 2

What will you decide—to keep yours or to exchange it for mine? (Circle one answer.)

(a) Keep
(b) Exchange

Treatment 4

What should you decide—to keep yours or to exchange it for mine? (Circle one answer.)

(a) Keep
(b) Exchange
(c) It doesn't matter

Thirty-three (33) students were asked the question from *Treatment 2* (*gut*) and 33 students were asked the *Treatment 4* (*rational*) question.

In the final treatment, *Treatment 5*, the first choice was removed altogether—there was no initial selection of an envelope. With this modification, the author intended to remove the endowment effect, anticipating that participants would not feel an allegiance to or ownership of a particular envelope. Without the endowment effect, the question suggested a rational probabilistic decision—thus only one scenario for this treatment. *Treatment 5* was answered by 33 students.

Treatment 5

I have two identical envelopes. You watch me put two $1 bills in one envelope and, in the other, I put a $1 bill and a $100 bill. We will create a situation where you may keep one envelope once I give you the choice.

Behind my back, I mix the envelopes and then place them on the table in front of you. We will call them envelope **A** and envelope **B**. I ask you to pick up envelope **A** and randomly select a bill from envelope **A** without looking. We see that the bill is a $1 bill. You return it to its envelope and place envelope **A** back on the table. You may now choose an envelope to keep, **A** or **B**.

Which envelope should you choose—**A** or **B**? (Circle one answer.)

(a) **A**
(b) **B**
(c) it doesn't matter

3 Results

Thirty-two (32) participants answered *Treatment 1* with their gut reaction and the majority (20) of those participants chose to stick with their original choice and keep their envelope—the "wrong" decision, while 12 chose to exchange their envelopes, which was the "correct" decision. This is consistent with, but not as extreme as, the percentages of individuals sticking with their first door in studies of the Monty Hall problem (Franco-Watkins et al. 2003; Petrocelli and Harris 2011). Perhaps Monty's non-random choice of which door to open does confound individuals' decisions. See Table 1.

In *Treatment 2*, there was an even greater tendency to stick with the initial choice, but in this case, sticking is the better probabilistic strategy. Twenty-five (25) out of the 33 participants who answered this question chose to stick with their initial choices. At the gut level, the endowment effect is strong—whether the participant's first choice is right or wrong. See Fig. 1.

In *Treatment 2*, the percentage opting to stick with their first choice is 76 % (versus 62.5 % in *Treatment 1*), but these participants have made the correct choice. The

Table 1 Survey results for *Treatments 1* through 4

	Treatment 1 First choice inconsistent with correct decision (Gut)	*Treatment 2* First choice consistent with correct decision (Gut)	*Treatment 3* First choice inconsistent with correct decision (Rational)	*Treatment 4* First choice consistent with correct decision (Rational)
Keep first choice envelope	20	25	13	15
Exchange envelope	12	8	8	11
It doesn't matter	–	–	11	7
n	32	33	32	33

Fig. 1 Gut reaction

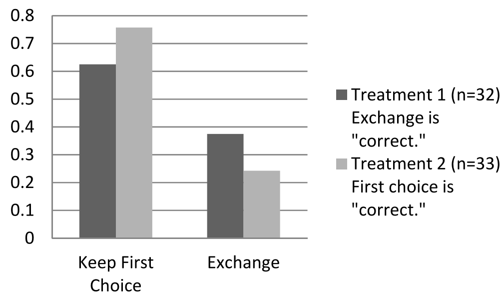
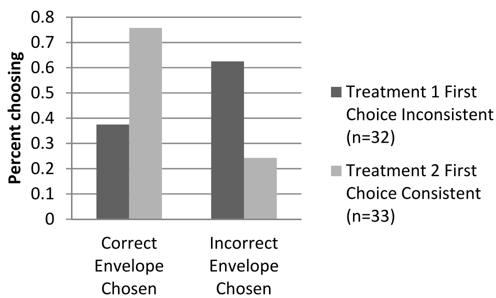


Fig. 2 First choice influence



author compared those making the correct choice when the first instinct is *inconsistent* with the correct answer (*Treatment 1*) versus those making the correct choice when the first instinct is *consistent* with the correct answer (*Treatment 2*). When the first instinct is inconsistent with the correct choice, the percent of correct choices versus incorrect choices is 37.5 % as compared with 62.5 %, but when the first instinct is consistent with the correct choice, the percent of correct choices versus incorrect choices is 76 % versus 24 % (see Fig. 2).

When one's first choice is consistent with clear probabilistic thinking we make correct decisions, but when the correct probability choices go against our instincts our gut will not let us give up our first choice. The author compared the numbers of

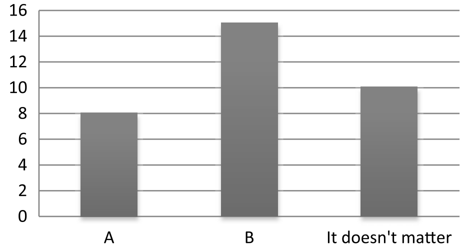


Fig. 3 Endowment effect removed

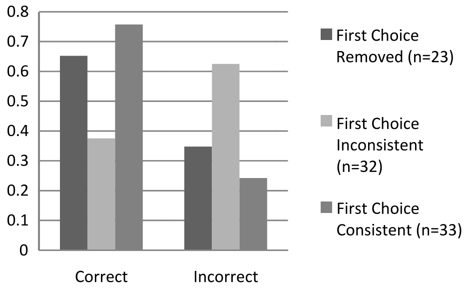


Fig. 4 Endowment effect comparison

correct and incorrect envelope choices in Fig. 2 to determine whether the distribution observed could have occurred at random. A two-way contingency table compared the counts of correct/incorrect decisions versus the treatment received, and the computed chi-squared statistic gave evidence that this distribution of correct/incorrect choices could not have occurred by chance (at the 0.02 level of significance).

Treatment 5 asked the participants to make a rational probabilistic decision without the endowment effect of initial ownership of an envelopment. Fifteen (15) out of 33 participants were able to correctly select envelope **B**. About half that number (8) incorrectly chose envelope **A**, and 10 participants said that it didn't matter which envelope was chosen (also incorrect). The results are illustrated in Fig. 3.

Assuming that those participants who responded that "It doesn't matter" which envelope was chosen would have chosen envelope **A** and **B** in equal numbers had they been forced to choose an envelope, the author removed those responses and compared the percentage of correct responses with the first choice removed with the gut responses from *Treatments 1* and *2*. Figure 4 presents those comparisons.

When the endowment effect is removed, 65 % of the participants can work out the correct choice versus 35 % who cannot. While this percentage of correct answers is not as high as when the first choice is consistent with correct thinking, in other words your *gut is right*, it does seem that participants can rationally and correctly evaluate the most promising envelope the majority of the time when they have no first choice investment. When the correct/incorrect percentages with the first choice removed are compared to those in the scenario of your gut interfering with correct choices, they are almost exactly reversed, 65 % correct and 35 % incorrect versus 37.5 % correct and 62.5 % incorrect. The participants are simply unable to override their gut instinct and correctly work out the probabilities.

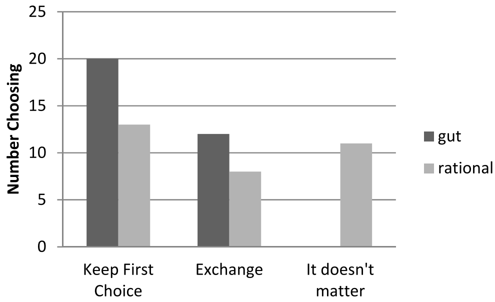


Fig. 5 *Treatments 1* and *3*

Treatments 3, *4*, and *5* attempted to determine whether a rational choice (System 2) can modify the endowment effect, overcoming the gut reaction to stick with one's first choice. Participants in *Treatment 1* and *3* were given the identical scenario—that of first choice inconsistent with the correct choice. The percentage choosing to stick with their original choice in *Treatment 3* was about 41 % versus the 62.5 % in *Treatment 1*, giving the appearance of the rational mind tempering the gut instinct. However, a smaller percentage opted to exchange—exchanging being the correct decision—25 % versus 37.5 %. A healthy 34 % decided that it didn't make any difference which envelope they chose. The hypothesis proposed in several other studies that most individuals believe the envelopes are equally likely to be the winner (as subjects have commented about Monty's unopened doors) is not fully supported. Figure 5 illustrates the comparison of gut versus rational.

Removing the 11 participants who declared that it didn't matter (whether to keep the first choice or exchange) from the results, results in 62 % sticking and 38 % switching in *Treatment 3*, remarkably similar to the 62.5 % sticking and 37.5 % switching in *Treatment 1*. The rational System 2 is not strong enough to overcome the first choice gut instinct.

Recall that in *Treatments 2* and *4*, the decision to stick with your first instinct is the correct choice. In the gut version (*Treatment 2*), 25 participants (76 %) decided to correctly keep their envelope. But when asked to make a rational choice, many participants indicated misgivings about that answer; only 15 participants (46 %) stuck with their initial choice. Seven participants (21 %) were persuaded by the more tempting, "It doesn't matter." The rational decision (System 2) can override an intuition, even when that intuition is correct. See Fig. 6.

Even removing the 7 participants in *Treatment 4* who indicated that it didn't matter which envelope was chosen, the gut versus rational choices are still at odds. In the gut version, 76 % correctly stick with their initial choice versus 24 % who exchange, while in the rational version 58 % stick and a whopping 42 % will incorrectly exchange.

The next comparison is among the scenarios that encouraged these participants to make a rational choice. *Treatment 3* asked the participants for a rational decision when their first choice was inconsistent with the correct choice; *Treatment 4* asked participants for a rational decision when their first choice was consistent with the correct decision; and *Treatment 5* asked for a rational decision with the first

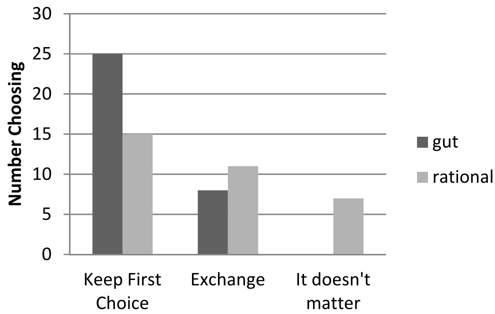


Fig. 6 *Treatments 2* and *4*

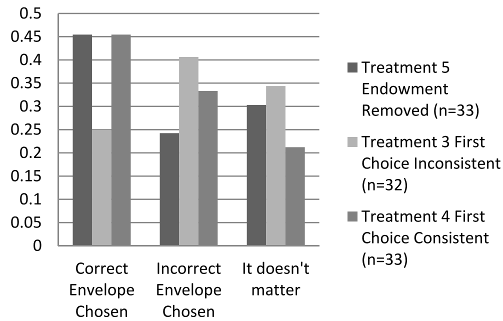


Fig. 7 Endowment effect versus rational choices

choice conflict removed. Figure 7 contrasts the responses from the participants in *Treatment 3, Treatment 4*, and *Treatment 5*. These were the rationally worded experiments where all of the participants were allowed to choose "It doesn't matter" (an incorrect answer in each treatment).

The participants in *Treatment 4* have everything going for them—the participants are urged to consider their decision rationally and their first-choice envelope is the correct envelope. Their probabilistic reasoning is not in conflict with their gut. Yet *Treatment 5* participants do equally well in making a correct probabilistic decision—about 45 % in both groups. With the endowment effect removed, these participants will make a rational, correct decision. The participants in *Treatment 3*, while encouraged to make a rational decision, were unable to because their rational minds were in conflict their first instincts.

Removing the "it doesn't matter" responses allows consideration of only the "committed" responses—the responses from those participants who were willing to commit to one envelope or the other. In these results, the *Treatment 5* participants did even better than the *Treatment 4* participants. An impressive 65 % chose the correct envelope. See Fig. 8.

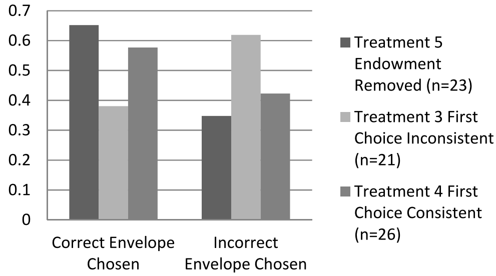


Fig. 8 Endowment effect versus rational choices (with committed choice)

4 Discussion

Franco-Watkins, Derks, and Doughery asked the question, "Why do people have the overwhelming propensity to stay with their initial selection [in the Monty Hall problem] even though the initial selection was chosen at random?" (Franco-Watkins et al. 2003, p. 88). Like Franco-Watkins, Derks, and Doughery, this author wondered why people insist on staying with their first choice. Previous research linked such behavior to the endowment effect coupled with the reluctance to tempt fate. Research also indicated that individuals prefer omission to commission and prefer to maintain the status quo; through inactions we do not initiate actions which tempt fate. And finally, the anticipation of the regret we will feel should tempting fate result in a negative outcome is "particularly likely to capture one's attention and imagination" (Risen and Gilovich 2008, p. 294). Tversky and Kahneman argued that the disproportionate availability of events in our memory distorts their perceived likelihood (Tversky and Kahneman 1973). The belief that it is bad luck to tempt fate (all the while believing that "there is no such thing as bad luck") is the result of two automatic mental processes: the tendency for thoughts and attention to be disproportionately drawn to the negative (loss aversion) and, having dominance in the imagination, the intuitive likelihood of such outcomes is enhanced. According to Petrocelli and Harris (2011), reflecting on the past is a critical component for profitable learning, but reflecting on what could have been (counterfactual thinking) can lead to dysfunctional behavior and memory distortion. Outcomes that are more vivid, outcomes that are more painful, and outcomes that we are more likely to regret are all overrepresented in our memories. Our mental sample spaces are distorted and unreliable.

Like the unwillingness of individuals to exchange a lottery ticket for an equally likely ticket even though the first ticket was dispensed randomly, I wondered if my participants would be unwilling to exchange a choice less likely to be correct for a choice more likely to be correct. This researcher asked, "Would the endowment effect dominate in a Monty Hall-type probability question where Monty wasn't able to rig the game?" The *Treatment 1* experiment most resembles the Monty Hall problem. In this experiment, 62.5 % wanted to stick with their initial choice and keep their original envelope versus 37.5 % who were willing to exchange and increase their chances of winning $101. As the endowment effect predicts, o*wning* the initial envelope makes it more difficult to switch.

Is there a difference in the way individuals evaluate a probability when their first instinct is inconsistent with correct probabilistic thinking versus when their first instinct is consistent with correct probabilistic thinking? To answer this question, the author compared *Treatment 1* and *Treatment 2*. Although there was a difference between the percentages sticking with their first choices, 62.5 % versus 76 %, all that can be said is that there was ample evidence of the endowment effect either way. When we are forced to commit to a choice (and given the freedom to go with our gut), the *ownership* of our first instinct trumps our rational decision.

The difference in the scenarios set up for *Treatment 1* and *Treatment 2* has implications for teachers of probability and statistics. The data indicate that flawed heuristics are very difficult to surmount; many are fraught with psychological baggage that gets in the way of correct thinking. If teachers ignore consequences of biases like the endowment effect, they can unknowingly steer students toward the wrong answers (by creating questions or problems where one's first choice is inconsistent with correct probabilistic thinking) or steer students toward the right answer (by creating questions or problems where one's first choice is consistent with the correct probabilistic thinking). In *Treatment 2*, 76 % reached the correct decision versus 24 % who reached the incorrect decision in *Treatment 1*. But of course, teachers won't want their students' decisions aided or impeded by the endowment effect; they will want their students to reach a correct probabilistic decision by understanding the mathematics behind the decision. They will want their student to think critically and not be guided by their gut instincts be they right or wrong.

Franco-Watkins, Derks, and Doughery, as well as many others, have proposed that there are two systems involved in choice and judgment: choice behavior is governed by the gut and probability judgments are governed by rules—sometimes the wrong rules (Franco-Watkins et al. 2003, p. 88). Can rational decision-making modify the gut instinct to stick with one's first choice? In other words, can System 2 monitor System 1 and override the endowment effect of owning one's first choice? *Treatment 3* and *Treatment 4* were administered as System 2 versions of *Treatment 1* and *Treatment 2* (the System 1 versions). A close examination reveals that the rational version may have moderated the endowment effect (fewer sticking with their first choice) but it did not promote correct probabilistic thinking. In fact, when asked to think rationally (in *Treatment 3*), fewer participants made the correct decision than in *Treatment 1*. Exchanging envelopes was the only correct decision and it was the decision made by 12 participants in the System 1 version and 8 participants in the System 2 version. The comparison between *Treatment 4* and *Treatment 2* was even more dramatic. When asked to think rationally, 15 participants made the correct decision versus 25 participants whose gut aligned with the correct decision. In this case, the rational decisions were curbing the (correct) gut instinct.

How can we encourage probabilistic thinking if our rational thought process cannot modify the endowment effect? Perhaps we should <u>remove</u> the initial ownership aspect of the problem. When participants were not allowed to initially choose an envelope (as in *Treatment 5*), it does appear that the majority of the participants can work out which envelope is more likely to be the winner. When committing to an envelope decision, 65 % of the participants reasoned out the correct envelope.

This author hypothesized that the unfounded belief that one should stick with one's first choice is one of the cognitive biases that inhibits probabilistic thinking. If superstition is an "irrational belief with regard to the unknown," then "refusing to switch," "sticking to your guns," or "going with your gut" might be considered a superstitious belief. Even when "new" information should inform our decision and encourage us to rethink our original decision, the influence of this commonly held (sometimes wrong) belief seems to trump probabilistic thinking.

5 Limitations

The questions designed for this study were original and the study has not been replicated. Replication of the study with a larger sample can bear out the results. The wording in the questions to address System 1 versus System 2 was an attempt to determine whether intuitive (gut) decisions are made differently from rational ones. Did the participants experience it as such? Did the participants understand the response, "It doesn't matter," to mean the chances were 50–50?[6] If they did understand the response to mean a 50–50 chance, there was no attempt to measure participants' equiprobability biases (see discussion by Watson 2005). In the treatment that attempted to remove the endowment effect by not allowing an initial envelope selection, participants were told to remove a bill from Envelope A and return it to the envelope. Can there be an endowment effect due to the *touching* of the envelope? If so, from the data the effect appears to be minimal. However, these questions linger and remain to be answered.

6 Conclusions

Granberg and Brown said, "Even when people have no good reason for initial selection, having acted on it, they become psychologically bound or committed to it" (as cited by Franco-Watkins et al. 2003, p. 88). The thought of exchanging our initial selection yields a conflict between our intuition and rational thought, and in everyday thought our likelihood assessments are a blend of both (Risen and Gilovich 2007). But these issues have implications far beyond Monty Hall games and lottery tickets. Any decision that involves making an initial choice, receiving updated information, and being allowed an opportunity to change your mind before making a final decision may be influenced by a flawed heuristic. Mathematics educators must become cognizant of students' preexisting probabilistic thinking before they try to teach them probability (Watson 2005). It is incumbent on teachers of probability to be aware that heuristics such as "sticking to your guns" exist and that the teachers

[6] Although the participants were instructed to select a multiple-choice answer only, two participants offered that their answer "It doesn't matter" was due to the fact that the odds were 50–50.

themselves can even be contributing to flawed reasoning by wording a problem in such a way that a gut reaction is summoned to quash rational probabilistic decision-making.

References

Benjamin, L. T., Cavell, T. A., & Shallenberger, W. R. (1984). Staying with initial answers on objective tests: Is it a myth? *Teaching of Psychology*, *11*(October), 133–141.

Edwards, K. A., & Marshall, C. (1977). First impressions on tests: some new findings. *Teaching of Psychology*, *4*(December), 193–195.

Franco-Watkins, A. M., Derks, P. L., & Dougherty, M. R. P. (2003). Reasoning in the Monty Hall problem: examining choice behavior and probability judgments. *Thinking & Reasoning*, *9*(1), 67–90. doi:10.1080/13546780244000114.

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *The American Psychologist*, *58*(9), 697–720. doi:10.1037/003-066X.58.9.697.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: the endowment effect, loss aversion, and status quo bias. *The Journal of Economic Perspectives*, *5*(Winter), 193–206.

Krauss, S., & Wang, X. T. (2003). The psychology of the Monty Hall problem: discovering psychological mechanisms for solving a tenacious brain teaser. *Journal of Experimental Psychology. General*, *132*(1), 3–22.

Kruger, J., Wirtz, D., & Miller, D. T. (2005). Counterfactual thinking and the first instinct fallacy. *Journal of Personality and Social Psychology*, *88*, 725–735. doi:10.1037/0022-3514.88.5.725.

Miller, D. T., & Taylor, B. R. (1995). Counterfactual thought, regret, and superstition: how to avoid kicking yourself. In N. J. Roese & J. M. Olson (Eds.), *What might have been: the social psychology of counterfactual thinking* (pp. 305–332). Mahwah: Erlbaum.

Petrocelli, J. V., & Harris, A. K. (2011). Learning inhibition in the Monty Hall problem: the role of dysfunctional counterfactual prescriptions. *Personality & Social Psychology Bulletin*, *20*(10), 1–15. doi:10.1177/0146167211410245.

Risen, J. L., & Gilovich, T. (2007). Another look at why people are reluctant to exchange lottery tickets. *Journal of Personality and Social Psychology*, *93*(1), 12–22. doi:10.1037/0022-3514.93.1.12.

Risen, J. L., & Gilovich, T. (2008). Why people are reluctant to tempt fate. *Journal of Personality and Social Psychology*, *95*(2), 293–307. doi:10.1037/0022-3514.95.2.293.

Ritov, I., & Baron, J. (1992). Status-quo and omission bias. *Journal of Risk and Uncertainty*, *5*, 49–61. Retrieved on July 25, 2011 from http://opim.wharton.upenn.edu/risk/downloads/archive/arch190.pdf.

Roese, N. J., & Olson, J. M. (1995). Counterfactual thinking: a critical overview. In N. J. Roese & J. M. Olson (Eds.), *What might have been: the social psychology of counterfactual thinking* (pp. 1–55). Mahwah: Erlbaum.

Skinner, N. F. (1983). Switching answers on multiple-choice questions: shrewdness or shibboleth? *Teaching of Psychology*, *10*(December), 220–222.

Tversky, A., & Kahneman, D. (1973). Availability: a heuristic for judging frequency and probability. *Cognitive Psychology*, *5*, 207–232.

Watson, J. (2005). The probabilistic reasoning of middle school students. In G. A. Jones (Ed.), *Mathematics education library: Vol. 40. Exploring probability in school* (pp. 145–170). New York: Springer.

Developing Probabilistic Thinking: What About People's Conceptions?

Annie Savard

Abstract Since the important work on reasoning under uncertainty by Kahneman and Tversky in the 1970s, the description of how people think about probability by using intuitions, conceptions and misconceptions have been studied in psychology and mathematics education. Over the years, the body of the literature have identified and studied many of them. Some conceptions, such as representativeness and availability, are well known. But not all of the conceptions have been studied many times and the conceptions presented in the literature usually don't relate them to each other. Therefore, it is now difficult to have a broader perspective on people conceptions of probability. In addition to that, some epistemological differences exist between the conceptions. Not all of them use the same kind of reasoning for addressing different aspects of probability. Thus, a broader perspective of people conceptions of probability involves not only knowing about conceptions and links them together; it also includes knowing about the mathematical aspect involved.

This chapter will define what a conception is and present a classification of some of them given in the literature, based on their epistemological differences.

1 Introduction

Once upon a time, in a grade 4 classroom, the students were doing experimentations with a five-colour spinner. They had to make a prediction before to spin and then record the outcomes. They were engaged in a classroom discussion about the comparison of the frequency of each colour gotten and thus about the variability of the results when a student said:

Felix: "It is not about that [*telling the number of times they got each colour*], *but about when we were said that we could choose a colour* [*on a spinner*]. *Me, I say that it can't always work. It is instead chance. Because I tried, I was on the purple* [*area*] *and I said that I would want to go on the red* [*area*], *but it did not work all the time. Because the hand spins fast and might go beside the red."*
Teacher: "Why do you want to go on the red area?"

A. Savard (✉)
McGill University, 3700, McTavish, Montreal, Québec, Canada H3A 1Y2
e-mail: annie.savard@mcgill.ca

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_15,

Felix: "Because I made a prediction. The predictions are not always true, it is not always true."

This student realized that the prediction of an event does not mean that this event will happen for sure. At the beginning of the lesson, he thought that predicting means that it will happen for sure, which is incorrect because making a prediction is reasoning under uncertainty. But then, realizing that it is not always the case by experiencing the variability of the results, he did not have the need anymore to use the prediction as a way to find out the result with certainty. He is now able to recognize that the outcomes depend of the randomness and thus he is now developing his reasoning under uncertainty using probabilistic reasoning (Savard 2010).

This little vignette or story is a good example of the challenge to think about probability. This phenomenon has been studied, and since the important work on reasoning under uncertainty by Kahneman and Tversky in the 1970s, the body of the literature have identified and studied how people think about probability by using intuitions, conceptions and misconceptions. This paper will present important works about students' thinking on probability and will classify them accordingly to their epistemological nature.

2 Thinking About Probability

The work done by Piaget and Inhelder in the early 1950s on the development of chance and randomness allowed many generations of researchers to know more about the construction of probability. Piaget and Inhelder (1974) studied the physical aspects of chance which is characterized by fortuitous events, i.e. irreversibility of mixtures and uniform or central distribution. They also studied the interpretation of chance given by children, which lead to the quantification of probability and, consequently, the construction of combinatorics. Their understanding of chance, in opposition of two types of causality, drove their studies. For them, chance is in opposition with mechanistic determinism and miracle. First, chance implies the irreversibility of spatial–temporal connections, which is in contradiction with the mechanistic determinism for which they are reversible. Second, chance suggests that it might have some physical laws or causality to explain a phenomenon, which is a contradiction with the notion of miracle. Thus, their work presented the ideas that an intuition of probability does exist and might be constructed. But the intuition of probability and the notion of chance can be considered as derived facts if they are "*compared with the search for order and its causes*" (San Martin 2007, p. 2). The search of causes or explanations might be seen as a way to make sense of a random phenomenon.

A large body of research on how judgement occurs in situation of uncertainty in the psychologist research and the mathematics educators fields emerged in the 1970s and 1980s (Shaughnessy 1992). Those researchers, among others Kaheman and Tversky (1972), Fischbein (1975), Falk (1983) and Konold (1989), observed and described how people think with probabilistic tasks. They qualified those judgements accordingly to their validity: these judgements reflected lacks and insufficiencies (Weil-Barais and Vergnaud 1990). Invalid judgements were named as primitive conceptions, cognitive illusions, misconceptions, fallacies in thinking, heuristics, judgemental biases or intuitions of probability. In fact, those judgement are considered errors of judgement and deviation of norms or rationality (Gigerenzer 2006). Intuitions, in particular, were considered in opposition with an analytic thinking, which is conscious, pure logical reasoning and independent of personal experience (Scholz 1991). Thus, intuitions and all kind of conceptions might have an implicit negative connotation. However, some psychology researchers[1] addressed the point by recognizing the rationality of those kinds of reasoning and the level of sophistication required. Those terms conception, preconception and misconception have been largely employed over the years by researchers in mathematics education. Furthermore, those terms conceptions and misconceptions are still largely used, without really being (re)defined. As a result, it seems that the use of those words still does reflect an invalid reasoning employed by students.

3 What Is a Conception?

Conceptions are explicative models for explaining the world (de Vecchi 1992; Giordan 1998; Vosniadou and Verschaffel 2004; White 1994). They also can be used for making predictions or for taking actions (Giordan 1998), for example, doing experiments. They are based on our daily life experiences (Vosniadou 1994) and therefore can be intuitive. In fact, conception is knowledge produced by the interaction between an individual and his/her milieu (Brousseau 1998; Gras and Totohasina 1995). A conception is based on personal experience and it is a mental filter to interpret a situation in order to make sense of it (Giordan and Pellaud 2004). Conceptions propose a simple and personal approach toward understanding a phenomenon or a problem arising from the environment. They are answers to regulations developed by a person to balance his/her cognitive structures at the time of the adaptation (Piaget 1967, 1974, 1975). They compensate the cognitive structures to face off a perturbation in the milieu. They are valid in certain contexts, but they cannot be generalized across all contexts. For this reason, they cannot be called misconceptions because they "worked" for some people in some contexts. As Giordan (1998) pointed out, a conception is not correct or incorrect, nor conform or inadequate, they are "operating" or inefficiency. In other words, they are useful or not for the person in the process of making sense, not for the teachers or other people. We prefer to call them *alternative conceptions* to reflect the validity of the conceptions in some contexts and their inadequacy when used outside their domain of validity. They are not seen as deviation of norms or rationality, nor as not logical, but rather knowledge in evolution. In fact, they are an alternative understanding. They can be personal,

[1] For instance, see the work of Gilovich and Griffin (2002) and Kahneman (2011).

created by one individual or socially shared, such as all conceptions identified by the scientific literature over the years.

A person's conceptions are influenced by the milieu and his or her experiences, history and projects. Thus, an alternative conception depends of the situation given, the task involved, the features and the wording of the situations (Bélanger 2008). So we can say that they are socially situated. They are close to common sense and to intuitive reasoning, but they are based on the person's explanation of the phenomena and therefore they present certain rationality, even when a magical thinking is being involved. Even if some aspects of the thinking such as magical thinking might be considered as irrational, there is rationality in the process of making sense of the world. In this sense, conceptions can't be pre-conceptions or spontaneous, because the organization of knowledge is inserted in time. The "starting point" of conceptions is hard to determine, because it changes from one person to another. It might be before or within formal or informal instruction. According to Taber (2000), a person may have incoherent and fragmentary conceptions or stable and internally coherent conceptions. In both cases, conceptions cannot be changed or replaced, they should be complexified (Désautels and Larochelle 1990; Larochelle and Désautels 1992). Because conceptions are not isolated inside de cognitive structure but instead integrated within, alternative conceptions are, in fact, part of a complex system (Bélanger 2008) or knowledge system (Vosniadou and Verschaffel 2004) and they are linked with other concepts as well (Caravita and Halldén 1994). The new knowledge does not destroy the older in this system; it is rather a reorganisation of the knowledge through interactions between the knowledge and the milieu which increase complexification of the organization (Désautels and Larochelle 1990; Larochelle and Désautels 1992). Alternative conceptions can be developed in parallel with the scientific knowledge (Savard 2008b). Thus, the stochastic thinking can be mixed with mythical or magical thinking (Steinbring 1991). According to Konold (1995) and Konold et al. (1993), different reasoning about uncertainty can be employed almost simultaneously in the same situation. For helping the students to learn, i.e. complexify their alternative conceptions, it is important to create a cognitive obstacle, where the learners should evaluate the validity of their alternative conceptions in a new context (Savard 2008a, 2008b). Then, they can choose other explanations for the phenomenon studied and thus do reorganise their knowledge. Doing so over the time might complexify their alternative conceptions.

4 Conceptions About Probability: Toward a Probabilistic Reasoning

Many conceptions about probability have been identified and studied by the literature over the years. Researchers expanded our knowledge about them in many directions. At this purpose, in 1992, Michael Shaughnessy presented a classification of conceptions. He defined four types of conceptions based on the conceptual understanding of probability and statistics, namely stochastics. He studied the reasoning

Table 1 Types of conceptions of stochastics (Shaughnessy 1992, p. 485)

Levels	Types of conceptions	Indicators
1	*Non-statistical*	Responses based on beliefs, deterministic models, causality, or single outcome expectations; no attention to or awareness of chance or random events.
2	*Naïve-statistical*	Use of judgemental heuristics, such as representativeness, availability, anchoring, balancing; mostly experientially based on non-normative responses; some understanding of chance and random events.
3	*Emergent-statistical*	Ability to apply normative models to simple problems; recognition that there is a difference between intuitive beliefs and a mathematical models; perhaps some training in probability and statistics; beginning to understand that there a multiple mathematical representations of chance, such as classical and frequentist.
4	*Pragmatic-statistical*	An in-depth understanding of mathematical models of chance (i.e. frequentist, classical, Bayesian); ability to compare and contrast various models of chance; ability to select and apply a normative model when confronted with choices under uncertainty; considerable training in stochastics; recognition of the limitations of and assumptions of various models.

involved in many well-known conceptions and associated a level of conceptual understanding of stochastics. Table 1 presents the types of conceptions of stochastics, according to Shaughnessy (1992).

To classify is interesting but presents some conflicting points of view with our definitions of conception which is based on our epistemological stance. All conflicting points of view shed light on what a conception is. The first conflicting point of view is about the value given to the conceptions. Conceptions are part of a person's knowledge and are valid in regard to the context used. For the person, the conceptions are valid; they are personal explanation in a given context. Classifying conceptions in a normative way accordingly to the conceptual understanding does not recognize the value given at the person's reasoning for making sense of the world. This classification is in opposition with our epistemology, that is, socio-constructivism. For the same reason, the second conflicting point of view is related to name given to the type of the conceptions. The name *Naïve-statistical* reflects a negative judgement on the ability to reason of a person. It is looks like a lack of knowledge instead of looking at the construction of knowledge; it does not reflect the process of learning. It is a label on conceptions showed at the moment.

The third conflicting point of view is about the role played by the conceptions in the learning process. The alternative conceptions can be developed in parallel with the standardized knowledge. So, a person might hold an alternative conception in a particular context and then have another answer in another or the same context. For example, it is possible for a student to explain why Heads or Tails is fair if the coin is not biased: “Both participants have equal chances to win, because the coin has two sides and we cannot say which side the coin will fall because it is not you who throws” (Savard 2008a, p. 4). In the same sentence, the student showed a correct representation of the mathematics involved in the situation and, in parallel, showed an alternative conception about the manipulation of the coin to explain the results. So, it is not possible to pretend that a certain type of conceptions might exactly explain the level of conceptual understanding of probability. For those reasons, we propose in this chapter another classification of conceptions of probability.

Alternative conceptions in probability are rooted in different epistemologies. Those epistemologies are underlined by the reasoning employed to think about probabilistic phenomenon. According to Konold (1989), reasoning about uncertainty involves two types of cognition: formal knowledge of probability theory and “natural assessments that become organized as judgements heuristics” (p. 88). We can translate that by the fact that, usually, people use a probabilistic reasoning or a deterministic reasoning to think about probability.

A probabilistic reasoning implies to reason under uncertainty. This reasoning takes in consideration two important components: the variability of the result and randomness. The result or outcome is not determined, it varies depending on the possible and favourable cases (theoretical probability) or the frequency (frequential probability) or some evaluation criteria (subjective probability) and it is related to randomness. Variability of the outcome might be represented by a distribution, on a table or on a graph (Canada 2006; Garfield and Ben-Zvi 2005). The outcome is randomly “selected”; it means that there is no correlation between the outcome and what’s happened before. In fact, randomness is uncertain, independent, without correlation, and it cannot be predicted with certainty (Dessart 1995; Dress 2004; Green 1993). For the teaching purpose, Batanero et al. (1998) defined the “essential characteristics of random phenomena:

1. In a given situation, there is more than one possible result.
2. The actual result which will occur is unpredictable.
3. There is a possibility—at least in the imagination—of repeating the experiment (or observation) many times.
4. The sequence of results obtained through repetition lacks a pattern that the subject could control or predict.
5. In this apparent disorder, a multitude of global regularities can be discovered, the most obvious being the stabilization of the relative frequencies of each possible result. This global regularity is the basis that allows us to study random phenomena using the theory of probability.” (p. 122)

In this sense, using a probabilistic reasoning takes into consideration variability and randomness. It implies also looking at the long-term results and not just the next outcome (Borovcnik and Peard 1996; Shaughnessy 1992). This way of thinking is different from other reasoning required by most school mathematics (Fischbein and Schnarch 1997), such as deterministic reasoning which is largely employed in sciences and mathematics (Moore 1990). Deterministic reasoning is not a feature of probability (Stohl 2005) and therefore is in contradiction with probabilistic thinking.

In fact, deterministic reasoning is a search for correlation, using present and past information to explain a phenomenon. There is dependence or cause between the events that might explain a result. The prediction has the meaning of exact prediction (Briand 2005), exactly what our little vignette has shown. According to Piaget and Inhelder (1974), the mechanistic determinism allows reversibility of spatial-temporal connections. Usually, there is no variability and uncertainty. The popular strategy of trial and error does reflect this reasoning: try a solution, get feedback and try again until you find the correct solution (Borovcnik and Peard 1996).

Those epistemological differences allow us to classify the alternatives conceptions about probability.

5 Methods

Our methodology consists in identifying the conceptions (also called heuristic biases, intuitions or misconceptions, and so on) in the scientific literature. We listed them out and for each we associated the authors who studied them. Then, we looked at the definitions and the examples proposed. In order to simplify the definitions, we wrote another one in our words. Finally, we looked at each conception and identified the kind of reasoning employed. The reader interested in knowing more about the tasks is encouraged to read, among others, the chapter written by Jane Watson (2005) in *Exploring Probability in School: A Challenge for Teaching and Learning*.

We present the conception into categories in Tables 2 and 3. Each table presents one category of conceptions about probability. Each conception is presented in alphabetical order and briefly defined. We cited the authors who published about them. This list of authors is not exhaustive, but it does reflect the trend about conceptions. Some conceptions might be sub-categorized. In this case, we present those under the name of the main category.

6 Classification of the Alternative Conceptions About Probability

The probabilistic conceptions are presented in Table 2:

Table 2 Probabilistic conceptions

Conception	Definition	Authors
Anchoring	Estimating based on initial values that were adjusted to the final answer. The initial value might be suggested by the words used in the problem or by a partial computation. Different starting points lead to different estimates which have initial values' bias.	(Tversky and Kaheman 1974) (Lecoutre et al. 1990) (Munisamy and Doraisamy 1998)

Table 2 (Continued)

Conception		Definition	Authors
	Conjunction fallacy	Considering that the conjunction between two events has higher probability than the one of them—overestimate	(Fischbein and Schnarch 1997) (Lecoutre and Fischbein 1998) (Scholz 1991) (Tversky and Kaheman 1974) (Zapata Cardona 2008)
	Disjunction fallacy	Considering that the conjunction between two events has higher probability than the one of them—underestimating	(Tversky and Kaheman 1974)
Effect on a sample size		When estimating, the size of the sample is not taken in consideration (law of large numbers).	(Fischbein and Schnarch 1997) (Lecoutre and Fischbein 1998) (Maury 1985) (Moore 1990) (Tversky and Kaheman 1971) (Zapata Cardona 2008)
Effect on the time axis (The Falk Phenomenon)		Conditional probability: inversion in the causality-effect axis. The outcome of an even comes after the conditioned event (Cause and effect).	(Falk 1983) (Fischbein and Schnarch 1997) (Lecoutre and Fischbein 1998)
	Cardinal conception	Interpreting the conditional probability $P(A\|B)$ as the ratio $\frac{CARD(A\cap B)}{Card(B)}$	(Gras and Totohasina 1995)
	Causal conception	Interpreting the conditional probability $P(A\|B)$ as an implicit relationship: B is the cause of and A is the consequence	(Gras and Totohasina 1995)
	Chronological conception	Interpreting the conditional probability $P(A\|B)$ as a temporal relationship: B should always precede A	(Gras and Totohasina 1995)

Table 2 (Continued)

Conception	Definition	Authors
Equiprobability	Considering that two events are equiprobable to occur when they are not Also called: Compound and single events	(Chiesi and Primi 2008) (Lecoutre and Durand 1988) (Lecoutre et al. 1990) (Lecoutre 1992) (Lecoutre and Fischbein 1998) (Fischbein et al. 1991) (Fischbein and Schnarch 1997) (Gürbüz and Birgin 2011) (Lecoutre and Fischbein 1998) (Konold et al. 1993) (Rouan 1991)
Outcome approach	Predicting on the next trial instead of considering the big picture	(Konold 1989, 1991) (Amir and Williams 1999) (Borovcnik and Peard 1996) (Haylock 2001) (Liu and Thompson 2007) (Watson et al. 1997)
Sample space	Not recognizing that all outcomes can occur	(Jones et al. 1997) (Chernoff and Zazkis 2011)
Visual effect or visual appearance	The segments of a spinner make different probabilities, depending on their place.	(Green 1989) (Jones et al. 1997) (Konold 1995) (Pratt Nicolson 2005)

The deterministic conceptions are presented in Table 3.

Table 3 Deterministic conceptions

Conception	Definition	Authors
Availability	Judgment is created on the availability of the information in the memory instead of the complete data.	(Amir and Williams (1999)) (Fischbein and Schnarch 1997) (Lecoutre and Fischbein 1998) (Garfield and Ahlgren 1988) (Konold 1991) (Munisamy and Doraisamy 1998)
Independence	The past results influence the future results. Results are linked to each other. Not recognizing randomness	Green 1990 (Maury 1985) (Moore 1990) (Pratt Nicolson 2005) (Watson et al. 1997)

Table 3 (Continued)

Conception	Definition	Authors
Personalist interpretation	Animism Attribution of phenomena to: God, luck or causality Almighty, fatality, good and evil The person: lucky, virtue, skills The milieu: conditions and objects in favour or not The rules of the game: compensation or amplification laws No rules or God not interested The stake (probability changes in regard to the stake) The mechanism of the object Manipulation of the device (technique of flipping)	(Truran 1995) (Konold 1989) (Amir and Williams 1999) (Brousseau 2005) (Garfield and Ben-Zvi 2005) (Kissane and Kemp 2010) (Maury 1985) (Savard 2008a) (Savard 2008b) (Savard 2011) (Truran 1995) (Vahey 2000) (Watson and Kelly 2004) (Watson and Moritz 2003) (Kuzmak and Gelman 1986) (Truran 1995) (Amir and Williams 1999) (Konold 1989) (Watson and Kelly 2004) (Watson and Moritz 2003)
Prediction	The prediction has the meaning of exact prediction	(Briand 2005) (Savard 2008b)
Representativeness	Events that seem to represent the chances better. The more representative an event is judged, the higher probability it receives. The population is similar to the sample.	(Kaheman and Tversky 1972) (Batanero et al. 1998) (Chernoff 2009) (Fischbein and Schnarch 1997) (Garfield and Ahlgren 1988) (Green 1983, 1990) (Gürbüz and Birgin 2011) (Konold 1995; Konold et al. 1993) (Lecoutre et al. 1990) (Lecoutre and Fischbein 1998) (Maury 1985) (Munisamy and Doraisamy 1998) (Saenz Castro 1998) (Zapata Cardona 2008)
Recency effects	Not taking into consideration the independence between the trials when observing the order of a sequence Wanting to balance them. Positive: same element of the sequence Negative: another element of the sequence (also named gambler's fallacy)	(Borovcnik and Peard 1996) (Brousseau 2005) (Chernoff 2009) (Chiesi and Primi 2008) (Fischbein et al. 1991) (Fischbein and Schnarch 1997) (Lecoutre and Fischbein 1998) (Munisamy and Doraisamy 1998) (Tarr and Jones 1997) (Zapata Cardona 2008)

Table 3 (Continued)

Conception	Definition	Authors
Unpredictability	Determining the result by the randomness or chance and thus not being able to evaluate or predict the probability	(Batanero and Serrano 1999) (Briand 2005) (Fischbein et al. 1991) (Lecoutre and Durand 1988) (Savard 2008b) (Savard 2010) (Thibault et al. 2012) (Vahey 2000)

7 General Considerations

It is interesting to note that some probabilistic conceptions are, in fact, related to combinatorics (English 2005). This is the case for *equiprobability* and *sample space*. For example, when students were asked to tell which event has greater chance of happening when rolling 2 six-sided dice (they had to choose between getting the pair 5–6, or the pair 6–6, or both have the same chance) it was more about permutations than about randomness. The students used a probabilistic reasoning to answer, but their answer did not take in consideration the permutations. For this reason, we consider this conception as a probabilistic conception because the thinking used is probabilistic.

According to Zapata Cardona (2008), researchers had explored *representativeness* under different names, that is, *conjunction fallacy, law of large numbers* and *gambler's fallacy*. We disagree with the first two claims, *conjunction fallacy* and *effect on a sample size* (*law of large numbers*), using a probabilistic reasoning can thus be seen as a deterministic conception. *Recency effect*, the *gambler's fallacy*, is a deterministic conception, another facet of *representativeness* (Tarr and Jones 1997).

The gambler psychologists have also studied the deterministic conception called *personalist interpretation* over the years. In 1975, Langer published an important article about the conception named *Illusion of control*. According to her, gamblers think that they have control of the game and thus try to develop some strategies to win. They expect their personal success to win to be higher than the real probability (Langer 1975). In fact, they attribute the success to themselves and blame the chance or other external factors for their failures (Langer and Roth 1975). In fact, *illusion of control* is giving an attribution to the results such as the *personalist interpretation*. Both use deterministic reasoning. In our point of view, the meaning is almost the same; the field of research differs. They are not in a schooling context, but rather in a gambling context, where adult, teenager and children cognition have been studied. But this difference should be taken in consideration. Of course, not only the research fields are slightly different, but also the contexts of the investigations give another interpretation of the person thinking: the thinking is oriented toward an important aim: wining a prize. In this sense, the affectivity here is higher for a person than being in a context of answering a question involving probability. This affectivity might occur when thinking about probability, because it is rooted in the personal experiences of life such as gambling or gaming activities. Our work on studying probabilistic reasoning in elementary school showed us that students might think about a probabilistic situation from a sociocultural point of view instead of using a probabilistic reasoning (Savard 2008a, 2008b). They were more concerned about the device, the luck or the way they manipulated the device than the mathematics involved in this particular situation. In this case, when some students showed a *personalist interpretation* or *an illusion of control*, they were not in a mathematical context: the sociocultural context was the milieu where they got their information about the situation. In this sense and in opposition with the other deterministic conceptions, *personalist interpretation* might not involve mathematical thinking at all. Whereas the *recency effects* might be rooted in a patterning reasoning and showing a deterministic explanation, the *personalist interpretation* might refer to a subjective point of view of fate or luck, without any mathematical rationale.

In the context of teaching and learning mathematics in school, we consider that this deterministic conception should be taken into consideration seriously in order for the mathematical reasoning to grow in every student.

References

Amir, G. S., & Williams, J. S. (1999). Cultural influences on children's probabilistic thinking. *Journal of Mathematical Behavior*, *18*(1), 85–107.

Batanero, C., Green, D. R., & Serrano, L. R. (1998). Randomness, its meaning and educational implications. *International Journal of Mathematical Education in Science and Technology*, *29*(1), 113–123.

Batanero, C., & Serrano, L. (1999). The meaning of randomness for secondary school students. *Journal for Research in Mathematics Education*, *30*(5), 558–567.

Bélanger, M. (2008). *Du changement conceptuel à la complexification conceptuelle dans l'apprentissage des sciences*. Montréal: Université de Montréal.

Borovcnik, M., & Peard, R. (1996). Probability. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematical education* (Vol. 2, pp. 239–287). Dordrecht: Kluwer Academic.

Briand, J. (2005). Une expérience statistique et une première approche des lois du hasard au lycée par une confrontation avec une machine simple. *Recherches en Didactique des sMathématiques*, *25*(2), 247–281.

Brousseau, G. (1998). *Théories des situations didactiques*. Grenoble: La pensée sauvage.

Brousseau, G. (2005). Situations fondamentales et processus génétiques de la statistique. In A. Mercier & C. Margolinas (Eds.), *Balises pour la Didactique des Mathématiques* (pp. 165–194). Grenoble: La pensée sauvage.

Canada, D. (2006). Elementary pre-service teachers' conceptions of variation in a probability context. *Statistics Education Research Journal*, *5*(1), 36–63.

Caravita, S., & Halldén, O. (1994). Re-framing the problem of conceptual change. *Learning and Instruction*, *4*(1), 89–111.

Chernoff, E. J. (2009). Sample space partitions: an investigative lens. *Journal of Mathematical Behavior*, *28*(1), 19–29.

Chernoff, E. J., & Zazkis, R. (2011). From personal to conventional probabilities: from sample set to sample space. *Educational Studies in Mathematics*, *77*, 15–33.

Chiesi, F., & Primi, C. (2008). *Primary school children's and college students' recency effects in a gaming situation*. Paper presented at the international congress on mathematics education (ICME 11), Monterrey, Mexico.

de Vecchi, G. (1992). *Aider les élèves à apprendre*. Paris: Hachette Éducation.

Désautels, J., & Larochelle, M. (1990). A constructivist pedagogical strategy: the epistemological disturbance (experiment and preliminary results). In D. E. Herget (Ed.), *More history and philosophy of science in science teaching* (pp. 236–257). Tallahassee: Florida State University.

Dessart, D. J. (1995). Randomness: a connection to reality. In P. A. House & A. F. Coxford (Eds.), *Connecting mathematics across the curriculum*. Reston: National Council of Teachers of Mathematics. 1995 Yearbook.

Dress, F. (2004). *Probabilités et statistiques de A à Z*. Paris: Dunod.

English, L. D. (2005). Combinatorics and the development of children's combinatorial reasoning. In G. A. Jones (Ed.), *Mathematics education library. Exploring probability in school: challenges for teaching and learning* (pp. 121–141). New York: Springer.

Falk, R. (1983). Revision of probabilities and the time axis. In D. Tall (Ed.), *Proceedings of the third international conference for the psychology of mathematics education* (pp. 64–66). Warwick: University of Warwick.

Fischbein, E. (1975). *The intuitive sources of probabilistic thinking in children*. Dordrecht: Reidel.

Fischbein, E., Sainati Nello, M., & Sciolis Marino, M. (1991). Factors affecting probabilistics judgements in children and adolescents. *Educational Studies in Mathematics*, *22*, 523–549.

Fischbein, E., & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal of Research in Mathematics Education*, *28*(1), 96–105.

Garfield, J., & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: implications for research. *Journal for Research in Mathematics Education*, *19*(1), 44–63.

Garfield, J., & Ben-Zvi, D. (2005). A framework for teaching and assessing reasoning about variability. *Statistics Education Research Journal*, *4*(1), 92–99.

Gigerenzer, G. (2006). Bounded and rational. In R. J. Stainton (Ed.), *Contemporary debates in cognitive science* (pp. 115–133). Oxford: Blackwall Publishing.

Gilovich, T., & Griffin, D. (2002). Introduction—heuristics and biaises: then and now. In D. G. D. K. T. Gilovich (Ed.), *Heuristics and biaises* (p. 855). Cambridge: Cambridge University Press.

Giordan, A. (1998). *Apprendre!* Paris: Belin.

Giordan, A., & Pellaud, F. (2004). La place des conceptions dans la médiation de la chimie. *L'actualité Chimique*, *280–281*, 49–52.

Gras, R., & Totohasina, A. (1995). Conceptions d'élèves sur la notion de probabilité conditionnelle révélées par une méthode d'analyse des données: implication—similarité—corrélation. *Educational Studies in Mathematics*, *28*, 337–363.

Green, D. (1983). From thumbtacks to inference. *School Science and Mathematics*, *83*(7), 541–551.

Green, D. (1989). School's pupil's understanding of randomness. In R. Morris (Ed.), *Studies mathematics education* (Vol. 7, pp. 27–39). Paris: Unesco.

Green, D. (1990). A longitudinal study of pupil's probability concepts. In D. Vere-Jones (Ed.), *Proceedings of the third international conference on teaching statistics* (Vol. 1). Dunedin: ISI Publications on Teaching Statistics.

Green, D. (1993). Ramdomness—a key concept. *International Journal of Mathematical Education in Science and Technology*, *24*(6), 897–905.

Gürbüz, R., & Birgin, O. (2011). The effects od computer-assisted teaching on remedying misconceptiosn: the case of the subject "probability". *Computers & Education*, *58*, 931–941.

Haylock, D. (2001). *Mathematics explained for primary teachers*. London: Chapman.

Jones, G. A., Langrall, C. W., Thornton, C. A., & Mogill, T. A. (1997). A framework for assessing and nurturing young children's thinking in probability. *Educational Studies in Mathematics*, *32*, 101–125.

Kaheman, D., & Tversky, A. (1972). Subjective probability: a judgement of representativeness. *Cognitive Psychology*, *3*, 439–454.

Kahneman, D. (2011). *Thinking fast and slow*. New York: Farra, Straus & Giroux.

Kissane, B., & Kemp, M. (2010). *Teachning and learning probability in the age of technology*. Paper presented at the 15th Asian technology conference on mathematics, Kuala Lumpyr, Malaysia, 17–21 December.

Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction*, *6*(1), 59–98.

Konold, C. (1991). Understanding student's beliefs about probability. In E. V. Glasersfeld (Ed.), *Radical constructivism in mathematics education* (pp. 139–156). Dordrecht: Kluwer Academic.

Konold, C. (1995). Issues in assessing conceptual understanding in probability and statistics. *Journal for Statistics Education*, *3*(1). http://www.amstat.org/publications/jse/v3n1/konold.html.

Konold, C., Pollatsek, A., Well, A., Lohmeier, J., & Lipson, A. (1993). Inconsistencies in students' reasoning about probability. *Journal for Research in Mathematics Education*, *24*(5), 392–414.

Kuzmak, S. D., & Gelman, R. (1986). Young children's understanding of random phenomena. *Child Development*, *57*, 559–566.

Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology*, *32*(2), 311–328.

Langer, E. J., & Roth, J. (1975). Heads I win, tails it's chance: the illusion of control as a function of the sequence of outcomes in a purely chance task. *Journal of Personality and Social Psychology*, *32*(6), 951–955.

Larochelle, M., & Désautels, J. (1992). *Autour de l'idée de science: itinéraires cognitifs d'étudiants et d'étudiantes*. Sainte-Foy: Presses de l'Université Laval.

Lecoutre, M.-P. (1992). Cognitive models and problems spaces in "purely random" situations. *Educational Studies in Mathematics*, *23*, 557–568.

Lecoutre, M.-P., & Durand, J.-L. (1988). Jugements probabilistes et modèles cognitifs: étude d'une situation aléatoire. *Educational Studies in Mathematics*, *19*, 357–368.

Lecoutre, M.-P., Durand, J.-L., & Cordier, J. (1990). A study of two biases in probabilistic judgements: representativeness and equiprobability. In J.-P. Caverni, J.-M. Fabre, & M. Gonzales (Eds.), *Cognitive biases* (pp. 563–575). Amsterdam: North-Holland.

Lecoutre, M.-P., & Fischbein, E. (1998). Évolution avec l'âge de « misconceptions » dans les intuitions probabilistes en France et en Israël. *Recherches en didactique des mathématiques*, *18*(3), 311–332.

Liu, Y., & Thompson, P. (2007). Teachers' understanding of probability. *Cognition and Instruction*, *25*(2–3), 113–160.

Maury, S. (1985). Influence de la question dans une épreuve relative à la notion d'indépendance. *Educational Studies in Mathematics*, *16*, 283–301.

Moore, D. S. (1990). Uncertainty. In L. A. Steen (Ed.), *On the shoulders of giants: new approaches to numeracy* (pp. 95–137). Washington: National Academy Press.

Munisamy, S., & Doraisamy, L. (1998). Levels of understanding of probability concept among secondary school pupils. *International Journal of Mathematical Education in Science and Technology*, *29*(1), 39–45.

Piaget, J. (1967). *Biologie et connaissance*. Paris: Gallimard.

Piaget, J. (1974). *La prise de conscience*. Paris: Presses Universitaires de France.

Piaget, J. (1975). *L'équilibration des structures cognitives*. Paris: Presses Universitaires de France.

Piaget, J., & Inhelder, B. (1974). *La genèse de l'idée de hasard chez l'enfant* (deuxième édition). Paris: Presses Universitaires de France.

Pratt Nicolson, C. (2005). Is chance fair? One student's thoughts on probability. *Teaching Children Mathematics*, *12*(2), 83–89.

Rouan, O. (1991). *Hasard et probabilité: conceptions d'élèves de 16 à 18 ans* (p. 40). Montréal: Université du Québec à Montréal.

Saenz Castro, C. (1998). Teaching probability for conceptual change. *Educational Studies in Mathematics*, *35*, 233–254.

San Martin, E. (2007). *Piaget's viewpoint o the teaching of probability: a breaking-off with the traditional notion of chance?* Paper presented at the international conference on teaching statistics, Salvador, Bahia, Brazil.

Savard, A. (2008a). *From "real life" to mathematics: a way for improving mathematical learning*. Paper presented at the International Congress on Mathematical Education (ICME 11), Monterrey, Mexico.

Savard, A. (2008b). *Le développement d'une pensée critique envers les jeux de hasard et d'argent par l'enseignement des probabilités à l'école primaire: vers une prise de décision. Thèse inédite*. Québec: Université Laval.

Savard, A. (2010). *Simulating the risk without gambling: can student conceptions generate critical thinking about probability?* Paper presented at the international conference on teaching statistic (ICOTS 8), Ljubljana, Slovenia, July 5–9.

Savard, A. (2011). Teaching citizenship education through the mathematics course. In B. Sriraman & V. Freiman (Eds.), *Interdisciplinarity for the 21st century: proceedings of the 3rd international symposium on mathematics and its connections to arts and sciences (MACAS)* (pp. 127–137). Charlotte: Information Age.

Scholz, R. W. (1991). Psychological research in probabilistic understanding. In R. Kapadia & M. Borovcnik (Eds.), *Chance encount probability in education* (pp. 213–256). Dordrecht: Kluwer Academic.

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. 465–495). New York: Macmillan Co.

Steinbring, H. (1991). The concept of chance in everyday teaching: aspects of a social epistemology of mathematical knowledge. *Educational Studies in Mathematics*, *22*, 503–522.

Stohl, H. (2005). Probability in teacher education and development. In G. A. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (Vol. 40, pp. 345–366). New York: Springer.

Taber, K. S. (2000). Multiple frameworks? Evidence of manifold conceptions in individual cognitive structure. *International Journal of Science Education*, *22*(4), 399–417.

Tarr, J. E., & Jones, G. A. (1997). A framework for assessing middle school students' thinking in conditional probability and independence. *Mathematics Education Research Journal*, *9*, 35–59.

Thibault, M., Lajoie, C., & Savard, A. (2012). *Un processus parsemé de choix déterminants: les dessous d'un cadre conceptuel au centre d'un projet de maîtrise portant sur des conceptions d'élèves autour du hasard et des probabilités*. Paper presented at the Colloque sur la formation à la recherche en didactique des mathématiques, Montréal.

Truran, K. (1995). *Animism: a view of probability behaviour*. Paper presented at the eighteeth annual conference of the mathematics education reserach group of Australasia, Darwin, NT.

Tversky, A., & Kaheman, D. (1971). Belief in the law of the small numbers. *Psychological Bulletin*, *76*(2), 105–110.

Tversky, A., & Kaheman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, *185*, 1124–1131.

Vahey, P. (2000). Learning probability through the use of a collaborative, inquiry-based simulation environment. *Interactive Learning Research*, *11*(1), 51–84.

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction*, *4*(1), 45–69.

Vosniadou, S., & Verschaffel, L. (2004). Extending the conceptual change approach to mathematics learning and teaching. *Learning and Instruction*, *14*(5), 445–451.

Watson, J. (2005). The probabilistic reasoning of middle school students. In G. A. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 145–169). New York: Springer.

Watson, J. M., Collins, K. F., & Moritz, J. B. (1997). The development of chance measurement. *Mathematics Education Research Journal*, *9*, 60–82.

Watson, J. M., & Kelly, B. A. (2004). Expectation versus variation: student's decision making in a chance environment. *Canadian Journal of Science, Mathematics and Technology Education*, *4*(3), 371–396.

Watson, J. M., & Moritz, J. B. (2003). Fairness of dice: a longitudinal study of students' beliefs and strategies for making judgements. *Journal for Research in Mathematics Education*, *34*(4), 270.

Weil-Barais, A., & Vergnaud, G. (1990). Students' conceptions in physics and mathematics: biases and helps. In J.-M. F. M. G. J.-P. Caverni (Ed.), *Cognitive biases* (pp. 69–84). North-Holland: Elsevier.

White, R. T. (1994). Commentary: conceptual and conceptual change. *Learning and Instruction*, *4*(1), 117–121.

Zapata Cardona, L. (2008). *How do teachers deal with the heuristic of representativeness?* Paper presented at the international congress on mathematics education (ICME 11), Monterrey, Mexico.

Commentary on Perspective II: Psychology

Brian Greer

1 Introduction

In responding to this sample of work on probabilistic thinking, my aim is to highlight some overarching issues and ask questions intended to be provocative, in the best sense.

Within mathematics, probability is fascinating from many perspectives. Particularly intriguing is that probability had to wait until the seventeenth century to receive formal mathematical treatment despite its widespread manifestations across centuries and cultures in practices such as divination, gambling, and insurance. As expressed by Davis and Hersh (1986, p. 21), "the delay in the arrival of the theory of probability is one of the enigmas of modern science". Hacking (1975) discussed this enigma at some length and evaluated several of the proffered explanations. Finding none of these convincing, he suggested that the shift in Europe towards empirical investigation as the source of authority for knowledge might have been the triggering factor.

Probability remains a domain of mathematics full of unsettled questions, and thus an important counterexample to the conception of mathematics as timeless and universal. Further developments in the mathematics of uncertainty, such as fuzzy logic and possibility theory, continue to emerge. Hacking (1990, p. 1) declared that "the most decisive conceptual event of twentieth century physics has been the discovery that the world is not deterministic". Yet, in his account of chaos theory, Stewart (1989, p. 22) commented that "mathematicians are beginning to view order and chaos as two distinct manifestations of an underlying determinism".

Moreover, as documented by Hacking (1975, 1990) the development of probability from its origins in the second half of the seventeenth century has been inextricably linked with issues of statecraft and sociological conceptions of human nature and behavior.

B. Greer (✉)
Portland State University, Portland, OR, USA
e-mail: brian1060ne@yahoo.com

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_16,

In short, what is clear is that there are complex mathematical, philosophical, humanistic, and scientific issues regarding probability. Naturally, the same goes for our understanding of probabilistic cognition. Rather than a simplistic "yes or no" question, Meder and Gigerenzer (2014) speak of identifying and characterizing the circumstances under with people are capable of sound probabilistic thinking—particularly when supported by helpful representational tools.

In the remainder of the chapter, I first make some comments on research of the type represented here, then on the teaching of probability. Following that, I consider several perspectives: the interplay between biological and cultural evolution; the limited relevance of neuroscientific inputs; the non-probabilistic mathematical underpinnings of probabilistic thinking; the epistemology of probabilistic thinking; the role of intuition; and the utility of dual-process theory. I finish with comments about the societal implications of how people use and misuse probabilistic thinking.

2 Comments on Research

To the extent that research in this area is carried out by giving self-contained probabilistic posers "cold" to undergraduates and school students, ecological validity remains a concern, together with questions about how the participants construe the experimental situation. For most participants, in most experiments, there are at most minor consequences. On the other hand, there are serious consequences when doctors do not understand the probabilities associated with testing positive for a disease or when judges do not understand probabilistic arguments in relation to evidence.

Further, while a great deal of methodological attention is generally paid to the sampling of participants, less attention is given to the sampling of tasks, representations, and assessment items from their respective populations. For example, as cited by Brase et al. (2014), Way (2003) presented a description of stages of development of probabilistic reasoning based on research with only two, very specific, tasks.

I find strange the notion that there is some kind of "natural" development of probabilistic thinking that occurs independently of instruction and culturally situated experiences beyond formal instruction. For this reason, I question the utility of asking whether the frequentist interpretation is "developmentally prior and privileged" (Brase et al. 2014). Even if that was a question that could be empirically settled (which I doubt), it is still arguable, and I would so argue, that instruction in probability should from the outset attempt to deal with the complex relationships between frequentist and other interpretations of probability, in particular since there are situations that we would normally include within the domain of probability to which frequentist approaches are not applicable.

I would extend part of the above argument to development of mathematical cognition in general, namely that you don't have to go far into mathematics in terms of complexity before the educational and social history of the student become determining factors in how they will perform. In my opinion, relevant information on these factors is generally under-reported in psychological research on mathematical cognition. (Design experiments offer greater access to such crucial information.)

Having expressed some critical concerns, I should stress the particular contributions that experimental research can make towards improving mathematics education. Most broadly, research as represented in these chapters serves to demonstrate the complexity of probabilistic thinking and the ubiquity, and resistance to change, of deviations from what is normatively regarded as correct. Research has accumulated a bank of specific misunderstandings and related tasks (see, for example, the taxonomy presented by Savard 2014) that constitutes a resource for textbook authors, curriculum designers, test constructors, and so on. (Whether they benefit as much as they could and should is another matter.) Another major contribution, particularly well represented in these chapters, is the conducting of design experiments based on representational tools. (I find surprising that there was not more discussion of the potential of computers to provide dynamic and interactive representations.)

3 Teaching Probability

As stated by Ejersbo and Leron (2014), "the goals of researchers in cognitive psychology and in mathematics education are quite different." The former, they suggest, are "interested in constructing an understanding of how the mind works" while the latter want to know what can be done through education, whether within or out of school. (The assumptions underlying the phrase "how *the* mind works" (emphasis added) should be interrogated, but that is well beyond the scope of this chapter.) Several of the chapters, including that of Ejersbo and Leron, bridge this gap in priorities, in particular by exploring the power of representational tools.

Teaching probability is inherently hard for many reasons. Most obviously, the attribution of a probability to an event is a construction that cannot be convincingly demonstrated with simplicity in the same way that an arithmetical statement such as $2 + 3 = 5$ can be modeled by small, discrete, identical objects or marks on paper. Moreover, there is not the societal support for learning about probability that there is for learning arithmetic. And, according to Fischbein, the cultural bias towards deterministic thinking is reinforced by education so that "the child approaching adolescence is in the habit (inculcated by instruction in physics, chemistry, mathematics, and even history and geography) of seeking causal relationships which can justify univocal explanations" (1975, p. 73). The situation is exacerbated by educational policy makers who do not take seriously the point made by many experimenters and mathematical educators that laying the foundations for probabilistic thinking needs to start early (Meder and Gigerenzer 2014).

If teaching children about probability is hard, teaching people to teach children probability is arguably even harder. Again the situation is exacerbated by top-down curriculum imposition that does not take this proposition seriously. Surveying curricular developments in 13 European countries, Howson (1991, p. 26) commented that "a curriculum cannot be considered in isolation from the teaching force which must implement it" and referred to "optimistic and untested assumptions concerning the teaching of probability". I see no indication that the current situation within typical educational systems is substantially different.

4 Various Perspectives

4.1 Interplay Between Biological and Cultural Evolution

The area of probabilistic thinking and decision making affords a context for discussion of the relationship between biological and cultural forms of evolution. We can speculate on the forms of decision-making that developed in response to survival in an uncertain world, but what can we reasonably expect biological evolution to have selected for? Hardly performance on Monty Hall's Dilemma, except in the broad sense of selecting for the ability to learn from experience.

There are three main reasons why I am skeptical that consideration of the evolutionary biological underpinnings will yield much of value to understanding probabilistic thinking. (Let me stress that my knowledge of the subject is very limited, so these arguments may be taken as naïve objections to be taken into account.)

Most obviously, there are clear contrasts with the modern industrialized world. For example, just as counting implies an abstraction whereby the counted units are considered equivalent, the frequentist idea of repeated events relies on the simplifying assumption that those events are indeed equivalent. These kinds of equivalence are more appropriate in an age of mass production.

Secondly, the whole conception of probability as a conceptual field is rooted in its cultural construction, the late emergence of which, in a formal sense, has already been commented on (though the earlier development of concepts of probability in India, where certainly combinatorics were extensively analyzed, bears further exploration, Raju 2010). Most of the situations that are implicated in research on probabilistic reasoning are cultural constructions that Savage (1954, cited by Gigerenzer and Todd 2012, p. 492) termed "small worlds" in which all the necessary information is perfectly known (or, to put it another way, in which the situation maps unproblematically on to the formal structures of probability theory). As they point out (p. 492), "the big question here is whether the small world is a true microcosm of the real ... world" (at whatever stage of evolutionary/cultural history).

Thirdly, and as well documented in several of the chapters, probabilistic thinking is mediated by representational tools. Actually, in general, it is hard to overstate the extent to which mathematics is constituted through its representations. Accordingly, an evolutionary perspective must also look at the development of human representational competence (Kaput and Shaffer 2002).

4.2 Contributions from Neuroscience

In general, I am skeptical that the neurosciences have much to offer when it comes to understanding probabilistic reasoning or how to effectively teach probability to children. One reason for the skepticism is that the neuroscientific research, like that in Artificial Intelligence and in neural networks, is concentrated on the more elementary parts of mathematics. This point of view was thus expressed by Lakoff and Nunez (2000, p. 26):

> At present, we have some idea of which ... capacities are innate... But ... when compared to the huge edifice of mathematics it is almost nothing. Second ... to know what parts of the brain "light up" when certain tasks are performed is far from knowing the neural mechanisms by which those tasks are performed.

Thus, even for the purposes of understanding how cognition works, the contributions from neurosciences, in their present state, are limited and the prospects of deriving implications for teaching mathematics in general, and probability in particular, seem premature (compare the earlier enthusiasm for neural networks). Moreover, if and when detailed descriptions of the neural substrates of certain cognitive aspects become available, it is not clear how the connections might be made to complex symbolically mediated (probabilistic) thinking, the necessary goal of any educational effort. As Abrahamson (2014) argues, "learners do not articulate tacit judgment directly in mathematical form; rather, they objectify these presymbolic notions in semiotic means made available to them in the learning environment."

4.3 Role of Proportional Reasoning

Fischbein (1975, p. 79) argued that "it is necessary to distinguish between the *concept* of probability as an explicit, correct computation of odds and the *intuition* of probability as a subjective, global estimation of odds." Piaget conducted groundbreaking studies of children's understanding of chance, causality and its absence, irreversible processes, and so on (Piaget and Inhelder 1975). But, arguably, he assimilated probabilistic cognition within his overall theory by placing great stress on combinatorics and proportional reasoning, linked to the theorized stage of formal operations. In general, there are clear indications that a narrow emphasis in elementary school on establishing proportional reasoning has detrimental side-effects, as has been clearly established through a program of research by my colleagues at Leuven University (De Bock et al. 2007). The problem is that students get to the point where "everything is proportional." This phenomenon shows up in many branches of mathematics, including probability. So, for example, most students believe that since the probability of getting a six from one roll of a standard die is $1/6$, the probability of getting at least one six in two, three ... rolls is $2/6$, $3/6$.... Similar errors have been made throughout history by mathematicians and others (De Bock et al. 2007, p. 11).

Another way in which this problem manifests itself is when people state that 200 heads in 300 tosses of a coin is as likely as 2 heads in 3 tosses. We may surmise that several years of instruction on equal ratios and fractions, and on the importance of ratios in the frequentist interpretation of probability, contribute to this fallacy and it seems reasonable to recommend that instead discrimination training be included from an early age to understand when "2 out of 3" and "200 out of 300" are equivalent and when they are not (which relates to the diagnostic problem and the emphasis on looking at frequencies rather than proportions, as discussed in several chapters).

4.4 Epistemology of Chance

To my mind, the chapters collectively pay insufficient attention to the epistemology of chance, as opposed to the technical details of quantifying probabilities. It has often been pointed out that there is a general reluctance to deal with uncertainty and variability, and a tendency to slide into deterministic modes of thinking and expression. Fischbein (1975, p. 131) stated that "whatever does not conform to strict determinism, whatever is associated with uncertainty, surprise, or randomness is seen as being outside of the possibility of a consistent, rational, scientific explanation." It is every easy to observe cases of slippage back towards the safety of determinism. For example, when polls are reported during elections, the discourse about a change in a candidate's estimated support is almost always framed in terms of conjectured causal explanation rather than sampling error. It is alarming when Ejersbo and Leron (2014) state that, given a base rate of 1/1000 for a disease "we know that [out of 1000 people] one person will have the disease." In fact, the theoretical probability that one person will have the disease is somewhat less than 1/2. (It is true that the expected value (in the technical sense) is 1, but the expected value in this sense is not the value that you expect—a prime example of the often user-unfriendly terminology in the field.) Generally, in the discussion of the diagnostic problem in several chapters here, the representation in terms of frequencies rather than probabilities is clearly effective, but it should at least be acknowledged that the justification for this move is not at all straightforward.

To turn to a different aspect, a pervasive aspect of probabilistic thinking is the relationship between the individual and the mass. It is natural for people to consider themselves as special, and difficult to relate intuitions about probabilities that personally attach to them with an objective analysis. This pervasive phenomenon is often exploited by the gambling industry, as in the advertising slogan for the British National Lottery "It could be you." As pointed out by Van Dooren (2014), judgments relating to oneself as a very special person are often laced with affect, and several related effects are well teased out in the experimental manipulations reported by Bennett (2014) in investigating Monty's Dilemma.

The relationship between oneself as an individual and mass phenomena is extremely hard to grasp. It is something that the Seattle photographer Chris Jordan has tried to elucidate through dynamic imagery representing very large numbers (www.chrisjordan.com). In discussing the motivation behind his work, he asks "How can we feel empowered as individuals when we are just one of 6.8 billion...?" and points out that "our collective behavior is causing a multi-leveled catastrophe, but each one of us is so small that our own impact is abstract and infinitesimal" (Jordan 2012, p. 255).

4.5 Intuition/Dual-Process Model

The general idea of two types of processing, one fast and intuitive (Type 1 or System 1), the other deliberative and logical (Type 2 or System 2) has been around for some time in experimental psychology and mathematics education (Gillard et al. 2009). Evans (2007, p. 168) listed 11 variants of dual-process theory within cognitive psychology. Stanovich (2009a) suggests that Type 2 processing has two levels, namely the reflective (whereby the person realizes that the intuitive response may need over-riding) and the algorithmic (whereby the normatively correct response is computed). I will not pursue the question here, but I think it is legitimate to ask (Greer 2009b, p. 155) what the ontological status of the two/three systems?

Type 1 processing is clearly very closely related to intuition, a form of thought that is universally acknowledged as important in mathematics, and probability in particularly, but hard to define and explain—arguably, Fischbein's (1987) attempt is the most complete so far. Mathematics educators have talked about ways in which the reflective disposition might be called into play, using such terms as cognitive "critics" (Davis 1984) and "alarm devices" (Fischbein 1987, p. 40). In Greer (2009b, pp. 155–156) I suggest that the tendency to answer quickly without reflection may be exacerbated by characteristics of school mathematics instruction.

5 Probability in Relation to Social and Political Issues

Davis and Hersh (1986, p. 24) distinguish three aspects of probability theory:

1. Pure probability theory: this is mathematical, axiomatic, deductive...
2. Applied probability theory: this attempts to fit probabilistic models to real-world situations...
3. Applied probability at the bottom line. Here the idea is to make practical decisions...

> Aspect 3 isn't really mathematics at all. It is decision, policy-making, public or private, backed up by the mathematics of aspects 1 and 2. Textbooks in probability are good on points 1 and 2. They are notoriously poor on point 3. The passage from aspects 1 and 2 to aspect 3 is accompanied by art, cunning, experience, persuasion, misrepresentation, common sense, and a whole host of rhetorical, but non-mathematical devices.

The distinction made here is somewhat parallel to that made by Meder and Gigerenzer (2014) between "decision making under risk" and "decision making under uncertainty." Much of contemporary life is influenced by mathematical models (often including probabilistic elements) that format aspects of our lives, often without any control on our point, or even knowledge of their assumptions and effects. Furthermore, because of feedback effects, "when part of reality becomes modeled and remodeled, then this process also influences reality itself" (Skovsmose 2000, p. 5). (Think of the likelihood that reporting of multiple polling results prior to an election may influence the outcome of the election.)

In my judgment, school mathematics ought to prepare people better, in terms of both a critical disposition and a sense of agency, to understand the roles and purposes of modeling, and in particular the limitations of mathematical models. As with word problems, examples of probabilistic modeling presented in school mathematics predominantly imply that the mapping of the situation onto a mathematical formulation is unproblematic. However, as argued by Greer and Mukhopadhyay (2005, p. 310), it is typically the case that a probabilistic model bears varying interpretations and they comment that "such openness is challenging for teachers who have a conception of mathematics as yielding unique, exact answers". (See Davis and Hersh (1986, p. 24) for an analysis of a historical problem generally considered to have a normative solution but for which they argue many alternative models are possible.)

In my judgment, research on probabilistic thinking likewise generally underplays the messiness of probabilistic modeling. For example, in Chiesi and Primi (2014), it is stated that "15 blue and 15 green marbles have been put into the bag shown in the video and one ball has been pushed in the see-through part. It was done a few times and a sequence of 5 green marbles was obtained." Then the child was asked to respond to the prompt: "the next one is more likely to be ...." But why should the child assume that the simulation is truly random? Suppose that the child considers, with some subjective probability, that the simulation is rigged. A sequence of 5 green marbles will raise this subjective probability.

The element of interpretation inherent in modeling becomes clear when experts disagree. Of course, this disagreement is to be expected within a framework where the aim is to persuade rather than to seek and try to tell the truth (the ethics of the advertising industry). For example, the antagonistic principle of legal cases means that, typically, both sides can and do call witnesses in support. The balance of persuasion tends to be based on how convincing the experts are rather than the soundness of the arguments, particularly given the low standard of understanding of probability not only among the public in the jury but also among the judges on the bench. And yet court cases, which typically involve the revision of subjective estimates in accordance with evidence, would *prima facie* appear to be a case for the application of Bayesian statistics.

As a second example of clashing experts, consider the case of the research done by a team of epidemiologists (Burnham et al. 2006) to estimate excess deaths in Iraq resulting from the invasion by the US and its allies (Greer 2009a). Technical details of the methodology used were disputed by numerous statistical experts, some saying they were appropriate, others that they were flawed. (Politicians, pundits, and the press responded predictably in accordance with their political positions.)

As stated by Meder and Gigerenzer (2014) "Teaching statistical thinking should be an integral part of comprehensive education, to provide children and adults with the risk literacy needed to make better decisions in a changing and uncertain world." I would go further to urge that the social and political implications be also addressed. At a conference on modeling and applications, I encountered resistance when I suggested that if gambling is to be used as a context within which to teach probability, then the teaching should include analysis of the social effects of gambling. An example of doing just that was described by Nobre (1989).

6 Final Comments

From the comfort of the critic's armchair, I can indulge in a wish list for development of research and teaching in this area.

For teaching probability, I suggest that a long-term view is needed, and that laying the foundations should begin in elementary school (Meder and Gigerenzer 2014). Such early foundations are arguably necessary, in particular, to prevent the consolidation of a bias towards determinism. In the school setting, it would be appropriate to move on from quick "in-and-out" studies to design research and longitudinal studies. (Of considerable relevance to the work of Abrahamson (2014), is a uniquely long-term study of combinatorial reasoning following a cohort of students from first grade through high school and beyond, including a task about the number of distinquishable towers that can be built with four bricks in two colors, Maher et al. 2010.)

Beyond school, I suggest more ethnographic studies of people thinking probabilistically for real, under normal human constraints, and with real consequences (Gigerenzer and Todd 2012), whether in law courts, hospitals, or bridge tournaments. (Bridge offers a fascinating interplay between probabilities and human factors. An expert player knows the theoretical odds inside out, but integrates them with subtle cues from bidding, and even such aspects as a subjective estimate of the skill of the opponent. There are situations in bridge structurally similar to that of Monty's Dilemma, and bridge theory includes "the principle of restricted choice" to explain these situations.) And studies directed at improving better collective understanding within society would be welcome, including attempts to educate people in the media, and even politicians.

References

Abrahamson, D. (2014). Rethinking probability education: perceptual judgment as epistemic resource. In E. J. Chernoff & B. Sriraman (Eds.), *Advances in mathematics education. Probabilistic thinking*. Dordrecht: Springer. This volume.

Bennett, D. (2014). Sticking to your guns: a flawed heuristic for probabilistic decision-making. In E. J. Chernoff & B. Sriraman (Eds.), *Advances in mathematics education. Probabilistic thinking*. Dordrecht: Springer. This volume.

Brase, G. L., Martinie, S., & Castillo-Garsow, C. (2014). Intuitive conceptions of probability and the development of basic math skills. In E. J. Chernoff & B. Sriraman (Eds.), *Advances in mathematics education. Probabilistic thinking*. Dordrecht: Springer. This volume.

Burnham, G., Lafta, R., Doocy, S., & Roberts, L. (2006). Mortality after the 2003 invasion of Iraq: a cross-sectional cluster sample survey. *The Lancet*, *368*(9546), 1421–1428.

Chiesi, F., & Primi, C. (2014). The interplay among knowledge, cognitive abilities and thinking styles in probabilistic reasoning: a test of a model. In E. J. Chernoff & B. Sriraman (Eds.), *Advances in mathematics education. Probabilistic thinking*. Dordrecht: Springer. This volume.

Davis, R. B. (1984). *Learning mathematics: the cognitive science approach to mathematics education*. London: Croom Helm.

Davis, P. J., & Hersh, R. (1986). *Descartes' dream: the world according to mathematics*. Sussex: Harvester.

De Bock, D., Van Dooren, W., Janssens, D., & Verschaffel, L. (2007). *The illusion of linearity: from analysis to improvement*. New York: Springer.

Ejersbo, L. R., & Leron, U. (2014). Revisiting the medical diagnosis problem: reconciling intuitive and analytical thinking. In E. J. Chernoff & B. Sriraman (Eds.), *Advances in mathematics education*. *Probabilistic thinking*. Dordrecht: Springer. This volume.

Evans, J. St. B. T (2007). *Hypothetical thinking: dual processes in reasoning and judgement*. New York: Psychology Press.

Fischbein, E. (1975). *The intuitive sources of probabilistic thinking in children*. Dordrecht: Reidel.

Fischbein, E. (1987). *Intuition in science and mathematics: an educational approach*. Dordrecht: Reidel.

Gigerenzer, G., & Todd, P. M. (2012). *Ecological rationality: the normative study of heuristics*. Oxford: Oxford University Press.

Gillard, E., Van Dooren, W., Schaeken, W., & Verschaffel, L. (2009). Dual processes in the psychology of mathematics education and cognitive psychology. *Human Development*, *52*, 95–108.

Greer, B. (2009a). Estimating Iraqi deaths: a case study with implications for mathematics education. *ZDM Mathematics Education*, *41*, 105–116.

Greer, B. (2009b). Helping children develop mathematically. *Human Development*, *52*, 148–161.

Greer, B., & Mukhopadhyay, M. (2005). Teaching and learning the mathematics of uncertainty: historical, cultural, social, and political contexts. In G. A. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 297–324). New York: Springer.

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

Hacking, I. (1990). *The taming of chance*. Cambridge: Cambridge University Press.

Howson, A. G. (1991). *National curricula in mathematics*. Leicester: The Mathematical Association.

Jordan, C. (2012). Running the numbers: a conversation. In S. Mukhopadhyay & W.-M. Roth (Eds.), *Alternative forms of knowing (in) mathematics: celebrations of diversity of mathematical practices* (pp. 247–259).

Kaput, J. J., & Shaffer, D. W. (2002). On the development of human representational competence from an evolutionary point of view. In K. Gravemeijer, R. Lehrer, B. van Oers, & L. Verschaffel (Eds.), *Symbolizing, modeling and tool use in mathematics education* (pp. 277–293). Dordrecht: Kluwer.

Lakoff, G., & Nunez, R. E. (2000). *Where mathematics comes from: how the embodied mind brings mathematics into being*. New York: Basic Books.

Maher, C. A., Powell, A. B., & Uptegrove, E. (Eds.) (2010). *Combinatorics and reasoning: representing, justifying and building isomorphisms*. New York: Springer.

Meder, B., & Gigerenzer, G. (2014). Statistical thinking: no one left behind. In E. J. Chernoff & B. Sriraman (Eds.), *Advances in mathematics education*. *Probabilistic thinking*. Dordrecht: Springer. This volume.

Nobre, S. R. (1989). The ethnomathematics of the most popular lottery in Brazil: the "Animal Lottery". In C. Keitel, P. Damerow, A. Bishop, & P. Gerdes (Eds.), *Mathematics, education, and society* (pp. 175–177). Paris: UNESCO.

Piaget, J., & Inhelder, B. (1975). *The origin of the idea of chance in children*. London: Routledge and Kegan Paul. (L. Leake Jr., P. Burrell, & H. D. Fischbein, Trans. Original work published in 1951).

Raju, C. K. (2010). Probability in ancient India. In P. S. Bandyopadhyay & M. R. Forster (Eds.), *Handbook of the philosophy of science: Vol. 7. Philosophy of statistics* (pp. 1175–1195). New York: Elsevier.

Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley.

Savard, A. (2014). Developing probabilistic thinking: what about people's conceptions? In E. J. Chernoff & B. Sriraman (Eds.), *Advances in mathematics education*. *Probabilistic thinking*. Dordrecht: Springer. This volume.

Skovsmose, O. (2000). Aporism and critical mathematics education. *For the Learning of Mathematics*, *20*(1), 2–8.

Stanovich, K. E. (2009a). Distinguishing the reflective, algorithmic, and autonomous minds: is it time for a tri-process theory? In J. St. B. T. Evans & K. Frankish (Eds.), *In two minds: dual processes and beyond* (pp. 55–88). Oxford: Oxford University Press.

Stewart, I. (1989). *Does God play dice?* London: Penguin.

Van Dooren, W. (2014). Probabilistic thinking: analyses from a psychological perspective. In E. J. Chernoff & B. Sriraman (Eds.), *Advances in mathematics education*. *Probabilistic thinking*. Dordrecht: Springer. This volume.

Way, J. (2003). The development of young children's notions of probability. CERME 3, Italy.

A Brief Overview and Critique of Perspective II on Probabilistic and Statistical Reasoning

Richard Lesh and Bharath Sriraman

Abstract In this overview and critique of perspective II, we briefly focus on several ways that *Models & Modeling Perspectives* (*MMP*) can be used to provide a unifying theoretical framework for developing both: (a) a coherent summary of the kind of research reported in this book, and (b) a useful list of testable claims that appear to be priorities to investigate in future research emanating from the kind of studies reported here. In particular, MMP was developed to identify: (a) emerging new types of mathematical thinking that appear to be needed beyond mathematics classrooms—including in situations involving new sciences (e.g., social sciences, engineering) not traditionally emphasized in K-12 curriculum materials, (b) new ways to operationally define important achievements currently being ignored in both school testing programs and documents specifying standards for teaching and learning, and (c) new teaching and learning opportunities made possible by new model-development tools for students (Lesh and Doerr, Beyond constructivist: a models & modeling perspective on mathematics teaching, learning, and problems solving, Erlbaum, Hillsdale, 2003; Lesh, Models & Modeling in Mathematics Education, Monograph for International Journal for Mathematical Thinking & Learning, Erlbaum, Hillsdale, 2003b).

Taken as a whole, the research reported in this book provides a variety of important building blocks for a research agenda that is aimed at some of the most important-yet-neglected areas of mathematics education research. Furthermore, many of these results are described ways that have implications far beyond topics related to statistics and/or probability. For example, many of the authors in this section deal in insightful ways with interactions between *intuitive* and *formal-analytic* ways of thinking. But, theory development doesn't always occur by continually expanding current ways of thinking (Lesh and Sriraman 2010). Quite often, long-term advancements in

R. Lesh (✉)
Indiana University, Bloomington, IN, USA
e-mail: ralesh@indiana.edu

B. Sriraman
The University of Montana, Missoula, MT, USA

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_17,

theory development depend on the *rejection* of plausible hypotheses—rather than on their continual expansion and refinement. And, rejected plausible hypotheses have been notably missing in mathematics education research—especially in areas related to the "mathematical practices" identified in curriculum standards documents such as the USA's newest Common Core State Curriculum Standards (CCSC 2010), where almost none of the deeper or higher-level goals are "operationally defined" in ways that are measureable. ... *Make sense of problems. Reason abstractly. Construct viable arguments. Model with mathematics. Use appropriate tools strategically. Attend to precision. Look for structure. Look for regularity. ... What does it mean to "understand" any of these practices? And, how can the development of these understandings be documented and assessed—or facilitated?* ... Beyond giving names and vague descriptions of the preceding goals, *CCSC Standards* provide little more than a few examples and lists of facts, skills, and procedures which fall far short of providing useful "operational definitions" of the what it means to understand the relevant underlying concepts and conceptual systems are believed to be associated with these practices. So, another strong characteristic of much of the research reported in this book is that it lays the kinds of groundwork that are needed to formulate testable hypotheses which might or might not end up being rejected. For example, many of the studies reported in this book clarify the apparent nature of important changes that have been occurring in both the theoretical foundations and the practical applications of statistical and/or probabilistic reasoning. And, at the same time, many of these studies also clarify some important shortcomings of current states of knowledge about teaching, learning, and problem solving—in topic areas related to probability and statistics, as well as in other topic areas.

In contrast to the procedural-understandings that continue to dominate the mathematics education communities' conceptions of curriculum goals, MMP-based research recognizes that the development of many deeper or higher-order conceptual understandings involve mainly the development powerful, sharable, and reusable models for describing situations mathematically, e.g., by quantifying them, systematizing them, dimensionalizing them, coordinatizing them, or "mathematizing them" in other ways. And, these models are not facts; they not skills; and, they are not procedures. But, in MMP-based research, they are expected to be among the most important higher-order conceptual goals of the K-21 mathematics curriculum.

MMP also recognizes that the emergence of new technology-based modeling tools has been: (a) changing (and/or reducing) the kinds of procedural knowledge that are most important for students to develop, (b) changing (and increasing) the levels and types of meta-level mathematical thinking that are needed for success beyond school classrooms, and (c) changing (and increasing) the kinds of conceptual knowledge that are needed to describe or conceptualize situations mathematically—in forms so that new computational tools can be used. For example, in topic areas related to probability and/or statistics, MMP recognizes that graphics-oriented computational-modeling tools are rapidly replacing calculus-based analytic models for making sense of realistically complex problem solving situations. Consequently, because of the increasing power and usefulness of these computational models, MMP recognizes that enormous changes are occurring in both the theoretical foundations and the practical applications of statistics and probability. Examples will be given in the remaining sections of this paper.

1 Why Focus on Probability & Statistics?

On the day that we began to write this chapter, the front page of the *New York Times* included an article describing how and why professional baseball teams increasingly want their radio analysts to be fluent in a category of advanced statistics that they refer to collectively as sabermetrics (e.g., WAR, VORP, BABIP). And similarly, in the entertainment section of this same newspaper, the lead article describes the plot of the recent movie, *Moneyball*, which tells a true story about how data-based hiring practices led to a rags-to-riches transformation of the *2002 Oakland A's Baseball Team*—enabling them to change from chumps to champions during a single season. Furthermore, in other parts of this same newspaper—in sections ranging from business, to *science*, to *education*, to *homes* and *gardening*—many of the articles include graphs, diagrams, and tables of data which reflect fundamental changes in both: (a) the situations that statistical models are being used to describe or explain, and (b) the graphics-intensive and computational nature of the statistics being used.

Based on the preceding kinds of observations, it is clear that recent advances in *computational modeling* have led to enormous changes in the kinds of dynamic, complex, and continually adaptive systems that strongly impact the lives of ordinary people—as well as professional scientists, and people in fields which have traditionally emphasized the use of quantitative methods. Yet, in the entire K-21 mathematics curriculum, there is (arguably) no other course in which so many exceedingly bright students so completely fail to understand the most basic concepts and principles they are intended to learn. For example, even in university-level courses for graduate students who have already demonstrated strong records of achievement in a wide variety of academic disciplines, high percentages of these students emerge from their statistics courses with only the most rudimentary awareness of the fact that (a) *every statistical or probabilistic procedure is based on a model which "mathematizes"* (*i.e.*, *mathematically describes or conceptualizes*) *the quantities, relationships, patterns, actions, and interactions in the situation being analyzed,—and that* (*b*) *different models* (*or mathematizations*) *often lead to significantly different computational results*. Consequently, students who are taught traditional procedure-oriented conceptions of topics tend to imagine that the mastery of statistics and probability depends mostly on learning to execute procedures correctly—and fail to recognize that the results of these computations often depend strongly on the ways the situations are quantified, systematized, dimensionalized, coordinatized, or in other ways "mathematized" (or conceptualized or interpreted mathematically).

Research on probabilistic and statistical reasoning also is especially important for the advancement of theory in mathematics education because, in the entire K-21 mathematics curriculum, there are (arguably) no other content areas where the emergence of new computational modeling tools have so radically changed both the theoretical foundations and the range of practical applications of the subject. For example, concerning changes in underlying theoretical constructs, it has been known for many years that *Bayesian* and *Fisherian* approaches to statistics and probability often enable problem solvers to investigate situations in which calculus-based analytic methods (if they can be applied at all) often provide only crude oversimplifications of the situations being investigated. Furthermore, the kind of computational models that are emphasized in *Bayesian* and *Fisherian* statistics provide powerful descriptions of a wider range of situations than those that fit calculus-based analytic models; and, they also tend to be based on straightforward combinations of basic concepts in arithmetic, measurement, algebra, and geometry. So, many important "big ideas" in statistics and probability tend to become far more accessible to far younger students; and, at the same time, the straightforward simplicity of these graphics-oriented computational models often leads to an explosion of new kinds situations in which new types are mathematical reasoning are becoming increasingly important beyond school. However, in research investigating these possibilities, it will be important to operationally define what it means to "understand" these new levels and types of *conceptual knowledge and abilities* that students should develop. Yet, even among the studies reported in this book, things like *intuitive thinking* tend to be described in ways that emphasize only negative characteristics (such as the absence of conscious, analytic, formal thinking) and generally fail to emphasize levels or types of intuitive reasoning that might be important in their own right—and that might play important positive roles in the gradual development of progressively more powerful forms of thinking (Lesh et al. 2007).

2 What Kind of Research is Needed to Investigate the Preceding Issues?

To explain the suggestions and claims that were mentioned in the preceding section, the examples that will be given in this chapter have been taken from course-sized research studies which have focused on MMP-based conceptions of teaching, learning and problem solving (e.g., Lesh et al. 2010a). These studies have been conducted with students ranging from average ability middle school children, to college-level teacher education majors, to graduate students in education or other social sciences. And, the courses have emphasized Bayesian and Fisherian approaches to statistics and probability—and have used computational modeling tools such as *Fathom* and *TinkerPlots*. Another distinctive characteristic of these courses is that more than half of the students' time during class periods have focused on small-group problem solving activities known as *model-eliciting activities* (*MEAs*)—which are 30–90 minute simulations of "real life" problem solving activities where the products that problem solvers produce tend to be tools and artifacts that: (a) are powerful (in the specific situations described in a given problem), sharable (with other people), and reusable (in situations beyond the specific emphasized in the problem that was posed), (b) focus on the dozen-or-so "big ideas" is statistics and probability, and (c) emphasize the development of visualization and model development abilities associated these "big ideas". But, perhaps the most important characteristic of these courses is that they were designed primarily to provide a conceptually rich research site where researchers can directly observe, document, and assess the development many levels and types of higher-order *conceptual understandings* that are difficult to investigate in other contexts. They were not designed to be ideal instructional materials. Nonetheless, typically, when some type of "control group" has been used to assess the achievements of students in these courses, the "experimental class" has tended to outperform the "control groups" by margins that business executives refer to as *six-sigma differences* (Lesh et al. 2010a). That is, failures are eliminated; and, the lowest-achieving students in the "experimental classes" tended to score higher than the highest-achieving students in the "control groups".

One observation that helps explain why it is sensible for the preceding kinds of results should occur is because the final tests for these courses usually consisted of two equal sized parts. The first focused or procedural abilities of the type emphasized in traditional teaching of traditional topics; whereas, the second focused on students' models and modeling abilities—and on their abilities to identify and assess strengths and weaknesses of alternative models. Consequently, because students in procedure-oriented courses seldom spend much time developing or assessing models, they seldom develop deep understandings about the fact that each and every one of the computational procedures they are learning are based on a model that is at best a useful simplification of the situations they are investigating. Therefore, it is not surprising that these traditionally taught students tend to score exceedingly low on the half of tests that emphasize the development and assessment of alternative models (Doerr and Lesh 2010). A more surprising observation is that our studies also suggest that students who focus on *model development* and *model assessment* need not sacrifice procedural competency—and tend to score at least as well as their traditionally taught counterparts on procedure-oriented test questions. Nonetheless, to achieve such goals, the kind of research that MMP considers to be most important focuses on the following kinds of questions.

- What models are emerging as most important in "real life" situations outside of school?
- What does it mean to "understand" the preceding models?
- How do the preceding understandings develop?
- How can development be documented and assessed?
- And only later: How can development be facilitated? Lesh et al. (2007)

3 What Are Some of the Most Important Assumptions Underlying a *Models & Modeling Perspective* on Mathematics Teaching, Learning & Problem Solving?

Design principles for MEAs have been described in a number of past publications (e.g., Kelly and Lesh 2000); and, in fact, for the entire courses that we've described in the preceding section, resources have been published on *ProfRLesh*'s YouTube site (http://www.youtube.com/user/ProfRLesh) where other relevant curriculum resources also are available in accompanying Internet drop boxes. These other resources include lesson plans, homework activities, and tools to help teachers or researchers observe, assess, or guide students' work.

Relevant foundations for MMP also have been described in a number of past publications (e.g., Lesh and Doerr 2003; Lesh 2003a) where theoretical and epistemological roots can be traced to both cognitive and social constructivists (e.g., Piaget 1970; Vygotsky 1962; Sriraman and Lesh 2007) as well as to *American Pragmatists* (Lesh and Doerr 2003, Chap. 28). But, for the purposes of this book, MMP's most important assumptions are briefly described below.

MMP emphasizes the fact that mathematical knowledge consists of more than lists of declarative statements (*i.e., facts*) *or condition-action rules* (*i.e., skills & procedures*). *It also includes concepts and conceptual systems for describing, explaining, manipulating, or predicting the behaviors of situations where some important type of mathematical thinking is useful. And, in mathematics, the preceding kinds of conceptualization activities involve quantification, systematization, dimensionalization, coordinatization, and in general mathematization* (*or describing experiences mathematically*). *But, in mathematics and mature sciences, the preceding kinds of descriptions* (*or conceptualizations*) *are referred to as models; and, even ordinary people who are not professional mathematicians or scientists are expected to develop models for making sense of their everyday experiences—including experiences which involve some type of probabilistic or statistical reasoning. Therefore, even in elementary mathematics, MMP-based research expects models & modeling abilities to be among the most important achievements associated with deeper or higher-order "conceptual understandings" of most "big ideas" in mathematics. And, the meanings of these models are expected to develop along a variety of dimensions—such as concrete-abstract, specific-general, intuitive-formal, and so on. So, when modeling activities require students to externalize their models and modeling activities, they tend to automatically produce auditable trails of documentation that reveal important information about the nature of the students' interpretations of the relevant situations—and about how these interpretations evolved* (*from a first draft, to a second draft, and so on*). *Consequently, externalization tends to be and important and inherent aspect of modeling. So, the models that emerge often provide straightforward ways to operationally define* (*i.e., document and assess*) *what it means to "understand" many of the most important higher-order goals of mathematics education. Furthermore, model development activities also provide straightforward ways to help students develop powerful, sharable, and reusable models for making sense of experiences that involve probabilistic or statistical reasoning?*

A primary goal of MMP research is to develop powerful and useful models to explain students' and teachers' models and model development achievements. And, an entry-level definition of a *model* is that it is simply a system for describing or explaining other system for some specific purpose. Consequently, stories or metaphors

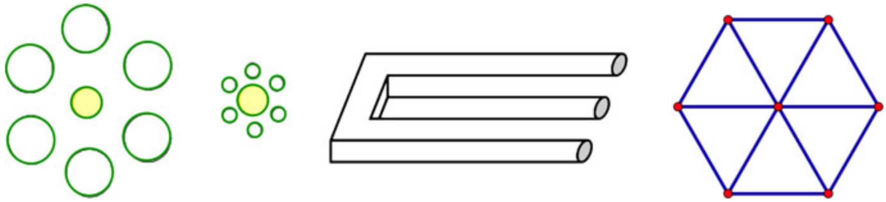


Fig. 1 Three simple perceptual illusions

can function as models; and, so can more formal and abstract structural metaphors. But, in any case, the system that is being modeled at one point in time may be used to model some other system at some later point in time. Also, sometimes, the preceding kinds of descriptive or explanatory systems are expected to function more like "windows" (that people look through in order to make sense of situations) rather than being treated as explicit objects of attention. So, people can think with models or they can think about them; and intuitive thinking can be thought about as occurring when problem solvers or decision makers think with (but not about) relevant models; whereas, formal-analytic thinking can be thought about as occurring when students think explicitly about relevant models.

The "perceptual illusions" shown in Fig. 1 are famous examples from *Gestalt Psychology*. They were developed to illustrate and emphasize ways that internal interpretation systems (i.e., gestalts) influence human perceptions. But, especially in early gestalt research, these "gestalts" (i.e., holistic interpretation frameworks) often were described in ways that suggested they resulted from hard wired patterns in human perception. Whereas later, such "gestalts" gradually came to be thought of in ways that modern researchers might think of as involving *conceptual illusions* more than *perceptual illusions*—or in ways that recognize that perception is a far more active structuring process than early perception theorists had imagined it to be. … Similarly, researchers such as Piaget investigated the development of the internal schemes (or cognitive structures) that human's develop to make sense of their experiences. And, most importantly, he investigated these cognitive structures developmentally. And, in many ways, MMP expects the development of students' models to be similar to the development of Piaget's cognitive structures—and to be quite different than hard-wired gestalts. On the other hand, in topic areas such as probability and statistics, MMP expects student-developed models to be significantly different than Piaget-like cognitive structures. For example, unlike Piaget, MMP researchers pay a great deal of attention to the roles of various representational media and tools play in the development of students' models; they investigate ways that student-developed models often integrate concepts and procedures from a variety of textbook topic areas or disciplines; and, their "subjects" often are teams of several students (or specialists representing diverse experiences and perspectives) rather than being just isolated individuals.

MMP research also investigate interactions among models from a variety of fields—rather than restricting attention to studies of single-topic decision-making. So, it becomes important to get clear about how mathematical models are similar-to yet-different-from models in other fields such as physics, chemistry, biology, or engineering. For example, in MMP-based research, models are considered to be "mathematical" (rather than physical, biological, or musical) if they focus on structural properties of the systems they are used to describe or explain. Similarly, one straightforward way that MMP-researchers distinguish one level or type of mathematical model from another is to look for: (a) different mathematical "objects" (e.g., quantities), or (b) different mathematical relationships, interactions, or transformations, or (c) different patterns or regularities which govern the behaviors of the preceding objects, relations, and interactions. And, one general conclusion from these developmental studies is that early stages of model development often involve the intuitive use of hodgepodges of undifferentiated and unintegrated conceptual systems (Borromeo-Ferri and Lesh 2011). So, in case studies of model development, models that end up sorting out and connecting ideas from several disciplines and topic areas seldom do so by putting together previously mastered single-topic models. In other words, MMP expects students to develop new models by gradually sorting our, integrating, and reorganizing fuzzy and overlapping interpretation systems, at least as much as they assemble or construct new models using previously mastered building blocks (Lesh and Yoon 2004).

Because *model-eliciting activities* (*MEAs*) have played a prominent role in MMP-based research, we should say a few more words about them here. *MEAs* are problem-solving or decision-making situations which are designed using principles which optimize the chances that significant types of model development will occur in observable forms during single 45 to 90 minute problem solving activities (Lesh and Zawojewski 2007). Also, in these model-development activities, the "products" that problem solvers produce usually involve more that just short "answers" to pre-mathematized questions. For example, the products often include the development of powerful, sharable, and reusable artifacts or tools (which can be thought of as being similar to spreadsheet with dynamic graphs that readers might develop to help the authors of this paper figure out their local, state, and federal taxes each year). So, the artifacts and tools that problem solvers develop during MEAs tend to include information about: (a) what mathematical "objects" or quantities are being considered, (b) what mathematical relationships or interactions are being considered, and (c) what patterns govern the behavior of the preceding objects, relations, and interactions. In other words, the products that students develop include explicit information about the nature of the models that they develop.

Why are the preceding kinds of assumptions so important in teaching & learning situations related to probability and statistics? One answer is that, in mathematics education, curriculum standards documents tend to emphasize two distinct types of objectives. The first type are declarative statements (facts) or condition-action rules (skills); and the second type include a variety of "higher order" objectives focusing on problem solving, modeling, or connections—or metacognitive understandings such as those related to beliefs, attitudes, dispositions, or thinking about thinking.

So, whereas the first type of objectives tend to be described in ways that are testable in straightforward and traditional ways, the second type are seldom operationally defined in ways that are measurable. So, even if these second types of objectives are described in ways so that teachers have some notions about what it might mean to emphasize them in their teaching, it is seldom clear how mastery of these objectives can be documented and assessed. So, in schools and school districts where testing programs are intended to strongly influence teaching and learning, these latter types of higher-order achievements tend to be ignored. Whereas MMP considers the development of powerful, sharable, and reusable models (and accompanying tools) to be among of the most important goals of mathematics education (Lesh and Lamon 1992).

4 What Changes Are Occurring in the Levels and Types of Probabilistic and Statistical Thinking That Are Important Outside of Mathematics Classrooms?

Clearly, answers to this question are not likely to emerge by asking only school teachers, teacher educators, and university professors in mathematics departments (i.e., the people whose opinions typically dominate curriculum standards documents). So, it is not surprising that even curriculum standards documents that claim to be future-oriented tend to be remarkably oblivious to things that have been happening outside the doors of school mathematics classrooms.

When MMP researchers first began to investigate the nature of "new kinds of mathematical thinking" that are becoming increasingly important outside of school, we often observed and interviewed people in fields like engineering or business management. But, in these studies, we became concerned because the things that we thought that we observed (or heard) seemed to be too strongly influenced by our own preconceived notions—as well by unreflective views of the people we observed or interviewed. So, we gradually migrated toward research methodologies that came to be called "evolving expert studies"—because all of the relevant participants (i.e., not only the people we observed or interviewed, but also people on our research staff) were required to express their opinions in forms that went through sequences of iterative express > test > revise cycles. And, one productive way of doing this was to engage both our research staff as well as "outside experts" (in fields such as engineering or business management) to collaborate with us to develop activities that represent high fidelity simulations "real life" situations where participants believed that "new kinds of mathematical thinking" could be observed, documented, and assessed.

Among other things, the preceding kinds of "evolving expert studies" led to the development of six principles that emerged as being especially important for investigating the nature of new levels and types of mathematical thinking that are becoming important in "real life" situations beyond school. But again, these six principles have been described and explained in a number of past publications (e.g., Lesh 2002;

Kelly et al. 2009). So, we will not repeat them in this chapter. However, a point that is relevant to mention about these six principles is that the people who developed them were surprised to find that nearly every problem in nearly every school textbook violates nearly every one of them. For example, in real life problem solving situations, the "products" that problem solvers produce often are not just short answers to pre-mathematized questions. They also may include the development of powerful, sharable, and reusable tools that can be used in a variety of decision making of specific situations. And, in such situations, the problem solver nearly always needs to know: (a) who needs the tool, and (b) what purposes does it need to serve. Otherwise, the problem solvers will have no basis for deciding such things as what level and type of product is most appropriate. Similarly, real life problem solving often involves trade-offs among factors such as costs and quality, or timeliness and durability. So, if no information is given about who needs the tool and why, then there is no way to respond to such issues (Hamilton et al. 2007).

In the preceding kinds of evolving expert studies, another observation that has emerged repeatedly is that the most important stages of model-development activities often involve quantifying qualitative attributes in ways that are useful. Consequently, in MMP-based courses investigating the development of students' statistical/probabilistic thinking, some important early activities engage students in identifying and classifying a variety of types of quantities that occur in everyday situations. Some examples are given in Fig. 2.

MMP research has found that, outside of school, one of the most common and important kinds of problems involves operationally defining some construct or attribute that cannot be measured directly (Zawojewski et al. 2009). For example, Fig. 3 shows a *Paper Airplane Problem* in which students are asked to develop a formula, or a rule, or a procedure, which judges in a contest can use to assess which paper airplane is the best "floater". This activity involves a *TinkerPlots* simulation of a *paper airplane contest* which MMP researchers designed to be similar to a problem they observed when a team of engineering graduate students were trying to operationally define the concept of "drag" for different kinds of moving objects.

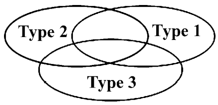
Solutions to this "floater" version of the *Paper Airplane Problem* involve a number of issues that occur quite commonly in a wide variety of real life situations. First, problem solvers not only need figure out a way to operationally define attributes like "floatiness" in a way that somehow uses weights, scales, and aggregates, or combinations of several qualitatively different kinds of information (e.g., distances, angles, times). And, they also need to combine information from several different throws. So, issues such as "centrality" and "spread" may need to be taken into account. Finally, the procedures that students develop also need to be quick and easy to use; and, the results that these tools produce also need to be consistent with people's intuitions (in cases that are easy to judge) while at the same time resolving problematic instances (in cases that are not easy to judge based on intuitions). So, trade-offs need to be considered between tools that are speedy to use and those that are highly accurate. And, in general, like many real life problems, but unlike most textbook problems, these *Paper Airplane Problems* involve problem solving under constraints—where trade-offs need to be considered.

YOUR TASK: During the next 60 minutes, your task is to create a classification system for sorting out examples of as many different kinds of quantities as you can find in a current newspaper, such as *USA Today*. Your classification system might look like a list, or a matrix, or a Venn Diagram; or, it might take some other form that you think is useful. But, in any case, it should help you and other students in this course recognize similarities and differences among a variety of different kinds of quantities that occur in everyday situations.

A List   A Matrix   A Venn Diagram

Type1
Type2
Type3
Type4
Type5
Type6

	Type1	Type2	Type3
Category1			
Category2			
Category3			
Category4			
Category5			



To help you get started on finding quantifiable attributes in your newspaper, here are a few examples that other students have found. So, you should be able to integrate them into your classification system.

- In Gymnasiums: distance, time (duration), speed, weight, difficulty, flexibility, fairness, intensity, strength, powerfulness, endurance, stamina, energy, boringness.
- In Counseling Clinics: empathy, friendliness, riskiness, anxiety, impulsivity, flexibility, bipolarity, masculinity, femininity, aggressiveness, forcefulness.
- In Kitchens: powerfulness (of a mixer), sharpness (of a knife), spottiness (of heat distribution in a pan or oven), costliness, temperature, heat, weight, volume, saltiness, smoothness, sweetness, doneness, freshness, crispness, difficulty, modifiability, reliability, fixability.
- In Music Studios: loudness, tempo, harmony, pitch, tone, clarity, style, melody, difficulty, powerfulness, creativity, modularity, popularity.
- In Photography Laboratories: sharpness, shadiness, distortedness, fixibility.
- In Science Laboratories: flexibility, power, energy, momentum, heat, temperature, velocity, acceleration, entropy, acidity, molarity, conductivity.
- In Ecological Systems: diversity, stability, interaction, entropy, robustness, power.
- In Weather Systems: rainy, sunny, windy, cloudy, humidity, low/high pressure, interaction.
- In Statistics: centrality, spread, distribution, interaction (correlation), probability, uncertainty, confidence, error, power, robustness.

Note: In this course, none of the preceding attributes will be referred to as "quantities" until they have been defined in ways that are measureable. Therefore, as you develop your classification scheme, it is important to notice that: (a) many of the quantifiable attributes cannot be measured directly, (b) many cannot be measured using standard units (like inches, seconds, degrees, and pounds), and (c) a variety of different kinds of scales might be needed to measure different attributes.

Quality Assessment Criteria: For the classification systems that students produced in this activity, several follow-up homework activities and whole-class discussions involved developing a scheme that can be used to assess the strengths and weaknesses of alternative classification systems. And, during these activities, the students googled several terms (such "intensive quantity") and recorded the following kinds of fact.

Fig. 2 Quantifying qualitative attributes

Attribute	How many? How much?	Of what?	For tasks or for people in a gym, what kind of units do you think would be sensible to use to measure the last three attributes in the third column?
Distance	3	Miles	
Time	30	Minutes	
Speed	60	Steps per Minute	
Weight	150	Pounds	
Difficulty	?	?	
Flexibility	?	?	
Strength	?	?	

A distinctive characteristic of *intensive quantities* is that they are *scalable* – but it is not sensible to *count* them. *Scalability* means that, if a whole quantity is broken up into parts, then the measure of each part is the same as the measure of the whole. For example, if the density of some materials is D, then the density of each of its parts also is D. So, if the density of one part is D, and the density of another part also is D, then the density of the two parts combined is not D+D. It is just D.

Intensive Quantities: The quantity being measured is essentially a relationship between two extensive quantities. Examples include: miles-per-hour, dollars-per-ticket, and other per, ness, or ity quantities (e.g., calories-per-minute, sweetness, density).
Extensive Quantities: The quantity being measured can be partitioned into units which can be added. Examples include: inches, apples, oranges.
Nominal Quantities: The quantity is assigned a name or an address that allows instances to be sorted out and compared in certain ways.

The students also found that measurements have two parts. One tells "how much" or "how many" or "what kind"; and the other tells "of what"—as well as what "unit" is being used to measure. And, answers to the question in the yellow box (on the right) are not restricted to "standard units"—such as inches, miles, pounds, or calories. They also may involve nonstandard units (like footsteps to measure distances). Furthermore, several different kinds of scales can be used in measurements that are made.

1. Nominal Scales often are used in ways that are like names on the jerseys of soccer players. But sometimes, they also may give additional information such as in addresses of houses: *1245 North Third Street Apartment 401*.
2. Ordinal Scales (e.g., 1st, 2nd, 3rd, *etc.*) specify an order for the items. But, distances between items are not specified. So, you cannot add or multiply measures. For example: $2 + 3$ is not necessarily the same as 5; and, 2×3 is not necessarily the same as 6.
3. Rational Scales. The zero point on the scale is significant; and, it usually refers to "none" of the relevant attribute. Furthermore, quantities can be added and multiplied.

Finally, the students noticed that many attributes or quantities they identified cannot be measured directly. Instead, they may be measured using formulas such as speed = miles-per-hour, or productivity = dollars-per-hour. In fact, in fields such as engineering and business management, some of the most common kinds of "mathematical thinking" that people do involve operationally defining attributes using some kind of formula. Examples include:

- *Aggregating qualitatively different kinds of information* to produce a single measure. For example, one such problem that is well known in MMP-based research gives students information about the results for try-outs at a baseball camp. The data involve: (a) running speed (in seconds), throwing distance (in feet), pitching accuracy (in inches), throwing speed (in miles-per-hour), consistent hitting ability (in hits-per-pitch), and power hitting ability (in feet—the average distance of the best three hits out of ten).

Fig. 2 (Continued)

- *Using ratios or rates to measure relationships between different kinds of information. For example, one such problem* that is well known in MMP-based research gives students information about the dollars-earned, and the time-spent-working each month (June–July–August) during different conditions for sales (busy, steady, and slow periods of sales). And students find a way to use this information in order to assess the productivity of workers (e.g., dollars-per-hour).

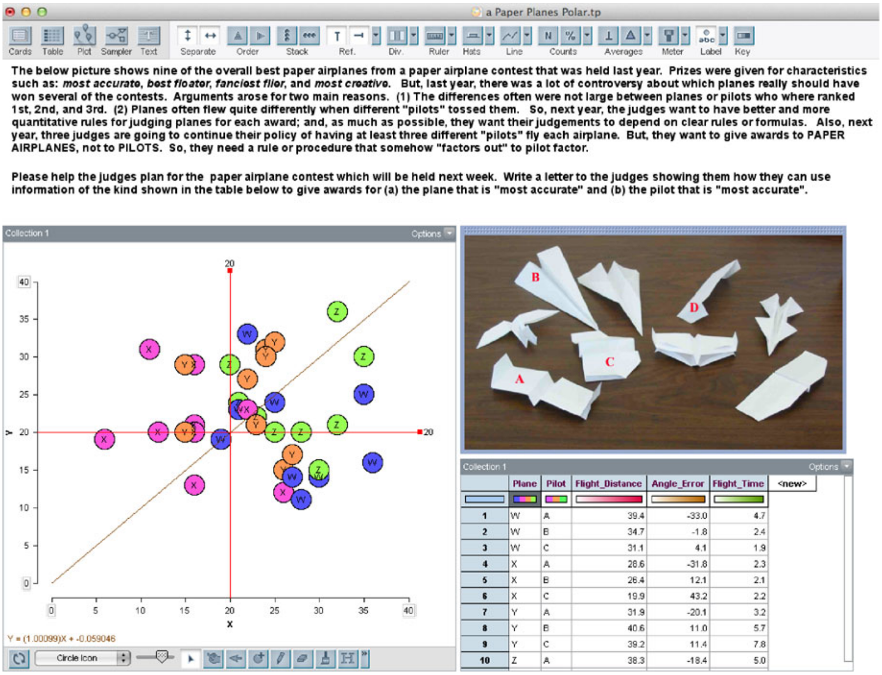
Follow-up activities to the newspaper problem also were designed to help students develop an awareness of the following kinds of issues.

- Whenever one kind of quantity (A) is used to determine another (B), issues may arise because "*A implies* $B1, B2, \ldots, Bn$" is not the same as "$B1, B2, \ldots, Bn$ *implies A*". That is, $B1, B2, \ldots, Bn$ may be "indicators" of A even if they do not operationally define A. For example, clocks are indicators of elapsed time. But, the reading on a clock can be changed without making any corresponding change in the time; and, similarly, thermometers can be used as indicators of temperature. But, the reading on the thermometer can be changed without making any corresponding change in the weather. Or, scores on a test of basic skills may be an indicator that higher-order achievements have developed, but efforts to increase test scores may not contribute anything to deeper understandings of important concepts in a given course.
- Another common type of measurement ability involves estimating the magnitude and types of errors that occur when measurements are made. For example, if a pedometer (i.e., an electric footstep counter) is used to measure distances between two locations in a park, errors can result from a variety of sources—including the length of walking strides and the choice of paths. And, such estimates are quite different than those typically occur in textbooks—where quick and informal calculations "in your head" are used to try to come close to results that would have occurred using slower and more careful calculations. … Another task that MMP researchers have investigated that emphasizes estimates of variability or error is a sling shot problem—where the students' task to figure out a reasonable answer to the following question. *How fast can a stone fly that has been shot by a modern hunting slingshot?*

Fig. 2 (Continued)

Figure 4 shows a *Darts Problem* that appears to be very similar to the *Paper Airplane Problem* in Fig. 3. And, these similarities are especially apparent if, in the *Paper Airplane Problem*, the landing positions of the planes are given using rectangular coordinates instead of polar coordinates. But, in the *Darts Problem*, another characteristic of "real life" problems becomes even more apparent. That is (and this is a point that engineers and other real life problem solvers continually emphasize): *Realistic solutions to realistically complex problems usually need to integrate ideas and procedures drawn from a variety of textbook topic areas* (*and disciplines*). For example, in the *Darts Problem*, the data are generated in a way that makes it obvious that a procedure for measuring "accuracy" is not likely to produce results that coincide with intuitive judgments (for easy-to-judge cases) unless it takes into account two different kinds of factors. One can be called *precision*; and, the other can be called *bias*.

One way to measure "precision" is to find the "center" of several throws and then to find how much these throws "spread" around this center. And, one way to



Notes:

1. The names of the four planes shown in the table were: (W) Wowzer, (X) X Wing Fighter, (Y) Yo Moma, and (Z) Zepher.
2. The names of the three pilots who tossed the planes were: (A) Abegail, (B) Beatrice, and (C) Cecelia.
3. The area of the gymnasium where the contest was held was 40 feet by 40 feet.
4. The planes were launched (i.e., tossed) from the corner of the floor corresponding to the lower left corner of the graph; and, the target that the planes were trying to hit was in the center of the floor corresponding to the point (20, 20) on the graph.
5. After each flight, the judges measured (i) the distance that the plane traveled from launch point to landing point, (ii) the time that the plane was in flight, and (iii) the angle that the flight path deviated from the diagonal line from the lower-left to the upper-right corners of the graph.

Fig. 3 The Great paper airplane contest

measure "bias" is to find how far the center of several throws is from the "bull's eye" in the center of the dart board. Consequently, even though rectangular coordinates are used instead of polar coordinates, it still remains true that two "dimensions" occur in the *Darts Problem* which are similar to the two dimensions that occur in the polar version of the *Paper Airplane Problem*. One dimension involves *precision*, and the other one involves *bias*. And, operational definitions of *precision* and *bias* both depend on how "centrality" and "spread" are defined for several throws. Then, in the *Darts Problem*, there are several sensible ways to define each of these constructs;

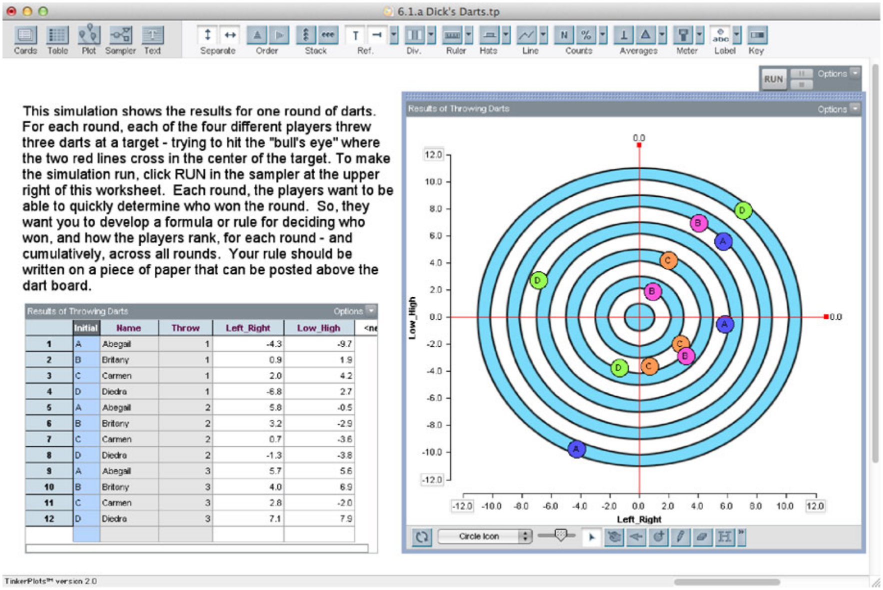


Fig. 4 Four players playing darts

and, different definitions often lead to different judgments about the overall accuracy of throws. In fact, in the Darts Problem, we generate the data in such a way that usual ways of thinking either centrality or spread do not resolve important cases. So, the judgments that are made do not coincide with intuitive judgments. Furthermore, many of these issues are highlighted in sequences of shorter follow-up problems that are aimed at acquainting students with the following kinds of facts.

- Even in simple situations like the three points at the vertices of a triangle, a variety of centers can be defined. For example, there are *incenters*, *circumcenters*, *centers of gravity*, *least distance points* (which minimize the sum of distances to a collection of points); and, *fractile centers* (which occur when a shape is projected into itself repeatedly). And, each of these notions of centrality is sensible under some conditions and for some purposes.
- For points on a line, there are *arithmetic means* ($A = (M + N)/2$) which also are called *centers of gravity*; there are *geometric means* ($G = \sqrt{}(M * N)$) which are useful in situations such as finding the average amount that money investments increase over several years; there are *harmonic means* ($H = 2/(1/M + 1/N)$ which apply to things like midpoints on musical scales); there are *root-mean-squares* ($R = \sqrt{}(M^2 + N^2)$) which are especially useful in situations that involve *Pythagorean* relationships among quantities; there are *fractile centers* which are points of convergence when a pattern is projected into itself repeatedly; and, there are *medians* (which also can be thought of as *least distance points* because they minimize the sum of the distances to the points that are given).

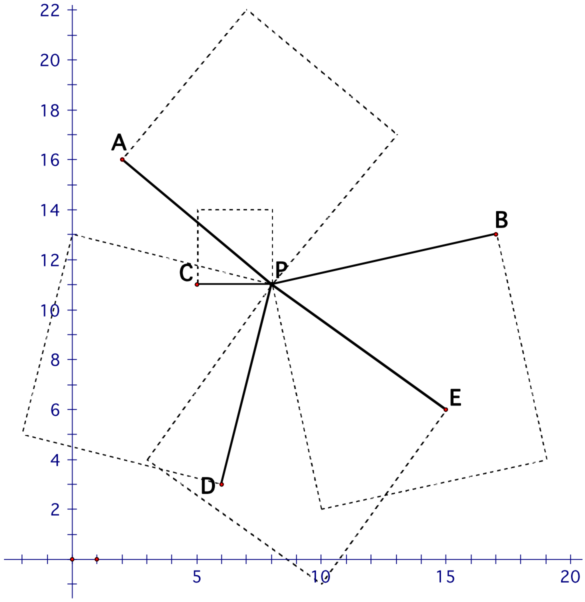


Fig. 5 Alternative conceptions of "centrality" for points on a line or plane

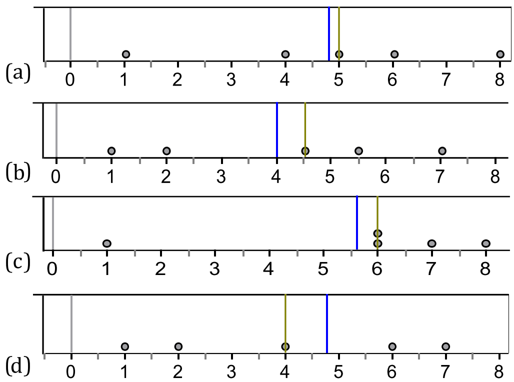
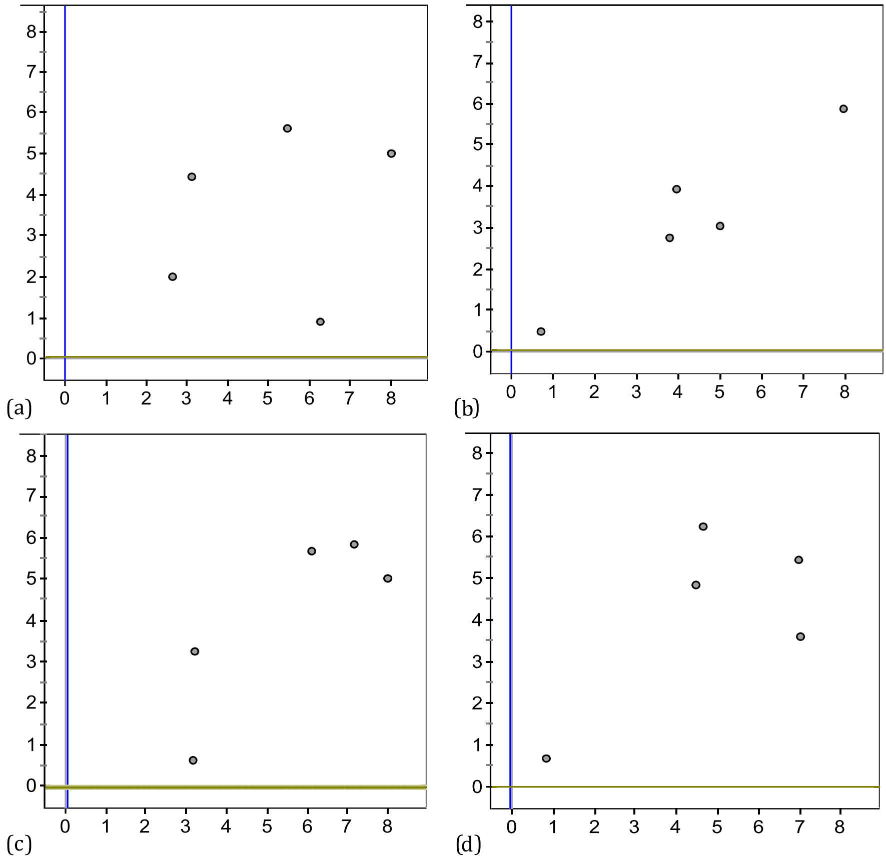


Fig. 6 Which of the following collections of points has the smallest amount of spread?

Figures 5, 6, and 7 describe some other important points that were emphasized in the homework activities that followed the in-class *Darts Problem*.

- For points in a plane (Fig. 5), the point that minimizes the sums of simple straight line distance is not the same as the point that minimizes the sums of the squares of the distances. Looking only at Fig. 5 students who work on the *Darts Problem* usually believe that the point that is most sensible to use as the "center" for several throws is the one that minimizes the sum of the distances to the three darts throws. But, they also usually imagine that this *least distance point* is the *arithmetic mean*. However, the *arithmetic mean* is the one that minimizes the squares of the distances to the given points. And, it gives the center-of-gravity of the given points.



Note: If attention shifts from *points of best fit* (i.e., centers) to *lines of best fit* or *curves of best fit* (such as those that are used in studies of correlation or regression), then similar kinds of issues arise as those described above. Furthermore, they also arise when attention shifts toward *hypothesis testing* (when “distances” between “centers” are measured) or *analysis of variance* (when variance is measured with respect to “centers”). In fact, slightly different notions of “distance”, “centrality”, and “spread” can strongly influence the results of analyses for each of the three basic types of problems that occur in introductory statistics courses, i.e., (a) comparing distributions of data for two or more collections, (b) comparing parameters for two or more collections of data, and (c) comparing relationships among two or more collections of data. And, such analyses also can be strongly influenced by the way information is quantified—and by the nature of scales that are used.

Fig. 7 Which of the following collections of points has the smallest amount of spread?

- For points shown in Figs. 6, 7, and 8, the *arithmetic mean* $(A = (M + N)/2)$ is not the point that minimizes the total sum of the distances to a collection of points (i.e., the *least distance point*). Instead, it is the *center of gravity*, and it minimizes

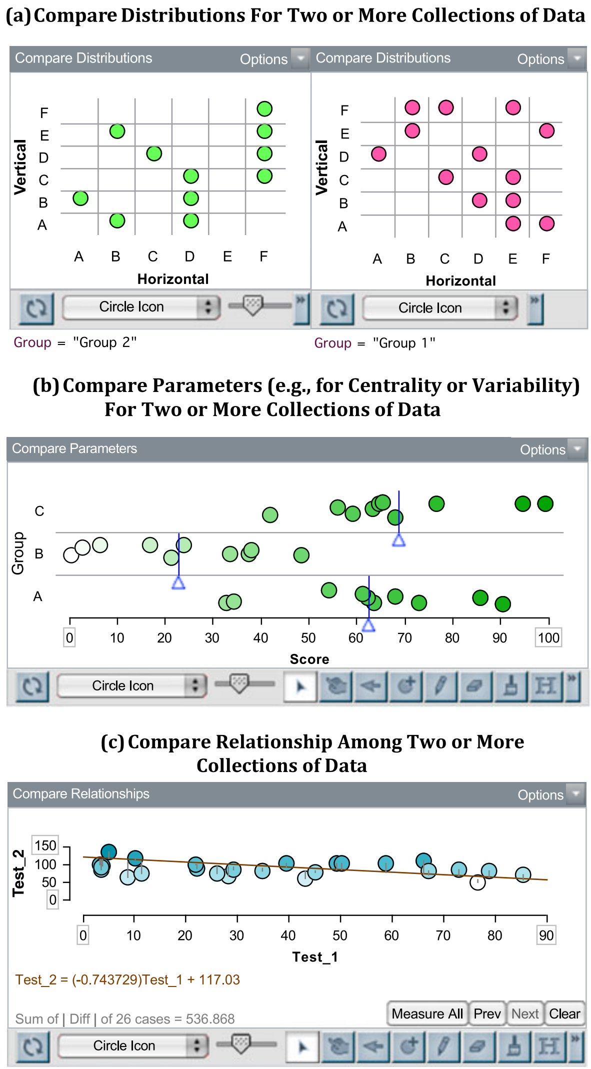


Fig. 8 Three basic types of problems in introductory statistics courses (Notice that all three types of problems depend on how distance, centrality, and spread are operationally defined)

the *sum of the squares* of the distances to the collection of points. So, it gives more weight to outlying points. . . . For points on a line, the *median* is the point that minimizes the sum of simple straight line distances. But, for points in a plane, the notion of a median becomes complex—because a point that is the median along one dimension is not necessarily the median along the second dimension. And, in either case, the location of the median depends on the location of the axes that are chosen. Furthermore, in some cases, more than a single point satisfies the criteria of being a *least distance point*. Nonetheless, using computational procedures it is easy to use a series of telescoping steps to choose a single point to be the *least distance point* for a collection of points.

In Figs. 6 and 7, the distributions that have the least and greatest "spread" depend on whether the mean or the median is chosen as the "center" of the distributions.

- In Fig. 6, the distribution with the least spread is (c) followed by (a), (b), and (d) if medians are used as the centers and if distances are measured using simple straight lines. But, the distribution with the least spread is (b) followed by (d), (a), and (c) if means are used as the centers and if distances are measured using sums of squares.
- In Fig. 7, the order from least to greatest spread shifts from (b), (d), (c), (a) to (a), (c), (b), (d) depending on whether "centrality" is defined as the sum of squares instead of the sum of simple straight line distances.

This section has focused on problem solving under constraints, problems whose solutions involve integrating ideas and procedures from more than a single discipline or textbook topic area, and problems that involve quantifying qualitative information, or "operationally defining" constructs that cannot be measured directly. And, many other equally important types of problems also could be mentioned. Primary among these would be problems that cannot be modeled using only a single one-way function (which also must be solvable and differentiable) with no feedback loops, no second-order effects, and consequently none of the characteristics of the kind of complex adaptive systems that are increasingly important in modern technology-based societies. Yet, complex systems often generate produces that are indistinguishable from those generated by probabilistic systems. So, understanding such systems involves many of the same issues as those that are important in statistics and probability. Furthermore, outside of school classrooms, situations where probabilistic and statistical thinking often occur are those where two or more agents interact. A sends something to B; B sends something to C; and, C sends something back to A. So, predictions and estimations play significantly different roles than they do in systems where feedback loops do not occur.

Many more examples could be given of types of problems and types of mathematical thinking that are important and occur often in "real life" situations—but that are largely ignored in school mathematics classrooms. But, our message here is simple. A large share of these problems and ways of thinking focus on models and modeling. But, the kind of modeling that we refer to here focuses on *mathematizing "real life" situations*—which is quite different than the way modeling is characterized in (for example) the USA's newest *CCSC Standards*. Whereas, MMP focuses on *mathematizing reality*, the CCSC Standards focus on realizing mathematics—or applying given ideas or procedures to textbook notions of reality (which typically fail to satisfy most of the six principles that apply to MMP-based *Model-Eliciting Activities*) (Lesh and Caylor 2007).

5 What Changes Are Occurring in Theoretical Models the Underlie Computational Approaches to Probabilistic and Statistical Thinking?

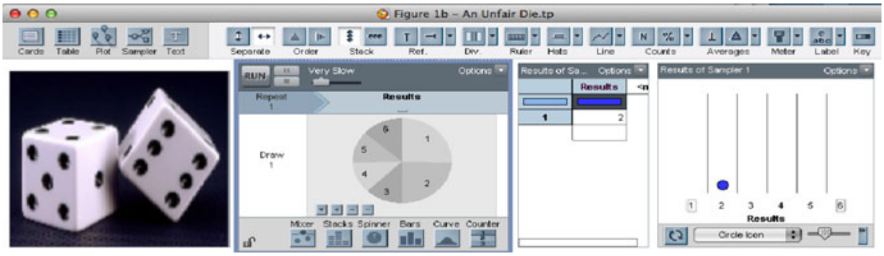
Among the problem situations described in the previous section, all involved descriptive statistics, not inferential statistics. Yet, descriptive statistics tend to be given little attention in introductory statistics courses. This practice, we believe, is unwise. And, one reason making this claim is that students who become skilled at developing mathematical models and simulations often find it quite straightforward and easy to develop their own data analysis procedures. Furthermore, the use of simulations lies at the heart of Bayesian and Fisherian approaches to statistics. So, this section will describe a variety of simulation-development activities; and, we will do so in situations that focus on probabilistic reasoning—and that are important in inferential statistics.

Once upon a time, it was a stroke of genius when an advisor to a medieval princes (who loved to gamble) when he figured out a way to "operationally define" the likelihood of an uncertain event. And, probability was born! The probability (P) of an event can be calculated by dividing the total number of ways that the desired outcome (D) can occur by the total number of possible outcomes (T), or $P = D/T$. But, the main factor that makes this simple formula difficult to think about is that it often is difficult to visualize (or count) either D, or T, or both. But, looking at Fig. 9(g), it's easy to see how simulations often make this formula quite easy to use. In fact, after creating such simulations, MMP researchers have found that even most middle school students are able to find probabilities for impressively complex events—such as finding the probability of winning in a game where four players each roll a pair of dice, and the winner is the one with the highest sum, or finding the probability that a given pair of dice is unfair.

Once the basic techniques have been mastered for the kind of dart simulations shown in Fig. 9, it is not a large step to more complex simulations like the one shown in Fig. 10. It show the outcomes from 1000 deals from a fair deck—where each deal includes five cards. To create this simulation, the most difficult task is to label the items in the sampler and record results is a way that is easy to organize and summarize. But, after this is achieved, it is fairly straightforward to answer questions such as: *What is the probability of being dealt a straight? Or a full house? Or a Flush?* ... It also is fairly straightforward to extend the simulation as shown in Fig. 10 so that it becomes the simulation shown in Fig. 11 which shows the result of dealing five-card hands to four different players. Then, it becomes a fairly simple process to extend the simulation so that the following kinds of questions can be answered. What are the chances of winning with a full house, if you are playing with three other players? Five other players? Six other players?

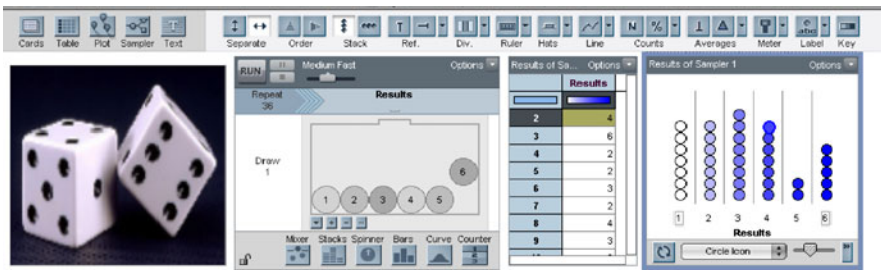
The point that all of these simulations demonstrate is that probabilistic reasoning becomes far less difficult when the results are both easy to visualize and count for both total number of desired results (or results of interest) and also the number of all total possibilities. Furthermore, such simulations also provide the foundations for

(a)
Figure 9(a) shows a TinkerPlots simulation of a single roll of a single cube. This simulation uses a spinner. But, others could use balls chosen from a bucket, or other random sampling devices.



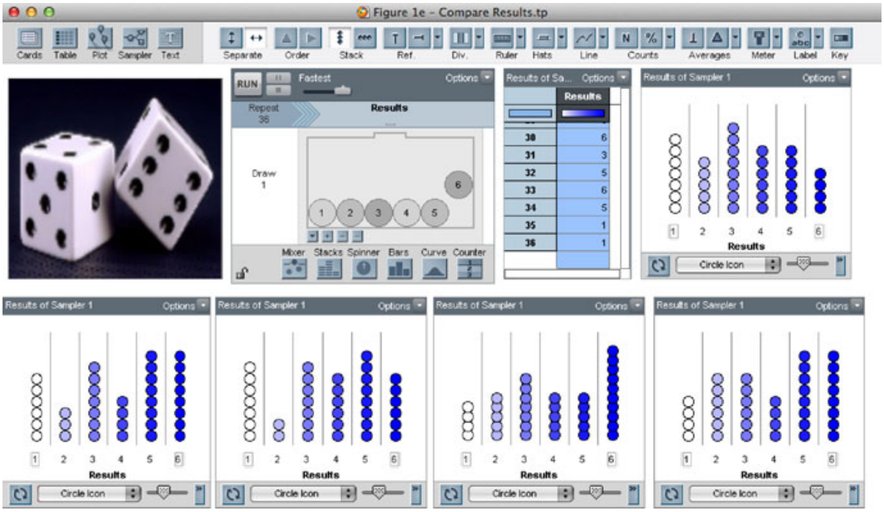
(b)
Figure 9(b) shows a simulation of a single roll of an unfair die. And, Fig. 9(c) shows the results from a single roll of a "mystery cube" (which might or might not be fair).

(c)

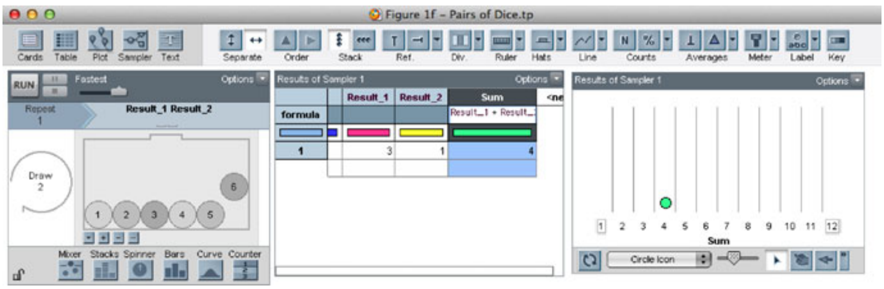


(d)
Figure 9(d) simulates rolling a single cube 36 times—and, recording how many results of each type 1–6.

Fig. 9 Simulations outcomes from rolling dice



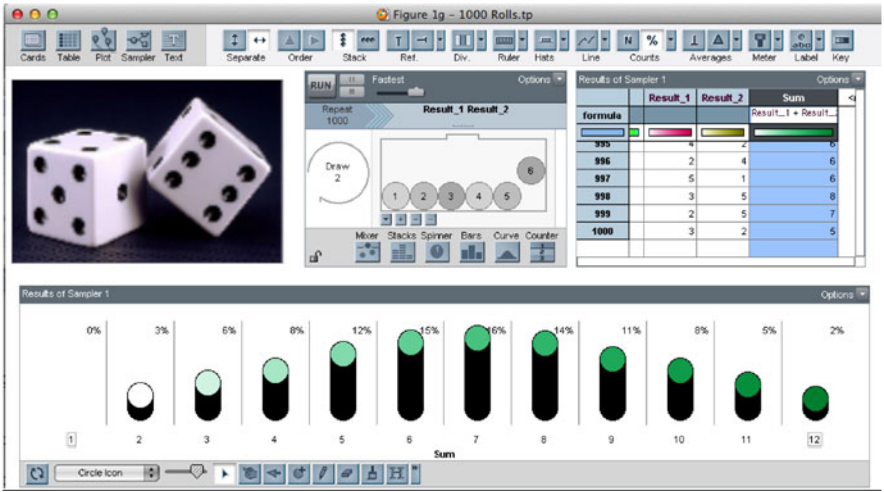
(e)
Figure 9(e) shows five different results from a from a "mystery sampler", and students are first asked to use only their intuitions to rank these results from least likely to most likely to have come from a fair sampler. Then, the students are asked to develop a rule which ranks each result from "most likely" to "least likely" to have come from a fair sampler.



(f)
Figure 9(f) shows simulates rolling a pair of fair dice, and recording the sum of the values on the pair.

Fig. 9 (Continued)

Bayesian and *Fisherian* approaches to statistical inference which depends on computational modeling rather than on analytic methods that attempt to describe data using functions such as normal distribution functions (and then trying to find ways to compensate for the fact that few real problems fit such functions exactly). So, computational modeling is changing the very theoretical foundations of probabilistic and statistical inference; and, it is doing it in ways that are both more powerful and easier to understand. In fact, once students have created insightful models of situations, they often discover that it isn't difficult at all to create their own statistical analyses procedures. For example, Figs. 12 and 13 show two problems that move in the direction of *analysis of variance* and *hypothesis testing*. And, both are again in-



(g)
Figure 9(g) shows the results from repeat the preceding procedure 1000 times. Then, students are asked questions such as: What is the probability of getting a 12 when rolling a pair of fair dice? What is the probability of getting a 7 or 11 when rolling a pair of dice? What is the probability of getting 10 or higher when rolling a pair of dice?

Fig. 9 (Continued)

formative examples about the nature of "real life" problem solving situations outside of school mathematics classrooms.

This brings us to the final category of research issues that we believe are priorities to investigate. These involve finding ways to help students develop deeper and higher-order understandings of the most important "big ideas" in statistics, probability, and other topic areas in mathematics. However, it is our firm belief that progress is not likely to be made on such issues using simple minded treatment-vs-control studies showing that "it works" in situations where the key characteristics of "it" are unclear and when "working" is measured using tests which do not measure the most important levels and types of understanding. On the other hand, progress is almost certain to occur through investigations that clarify: (a) the nature of emerging new types of "real world" problem solving, (b) what it means to "understand" relevant concepts and abilities in the preceding situations, (c) how such understandings develop, and (d) how development can be documented and assessed. But, ironically, to investigate such issues, we believe that it will be important for mathematics education researchers to shift attention beyond studies investigating isolated concepts, and shift attention toward the interacting development of larger collections of related concepts. And, this is especially true if attention shifts beyond the development of procedure-dominated achievements toward the development of powerful models and modeling abilities which focus on deeper and higher-order conceptual understandings. In other words, to investigate such issues, the kind of larger learning environments that will be needed are like the course-sized materials described in this chapter (Lesh et al. 2010b).

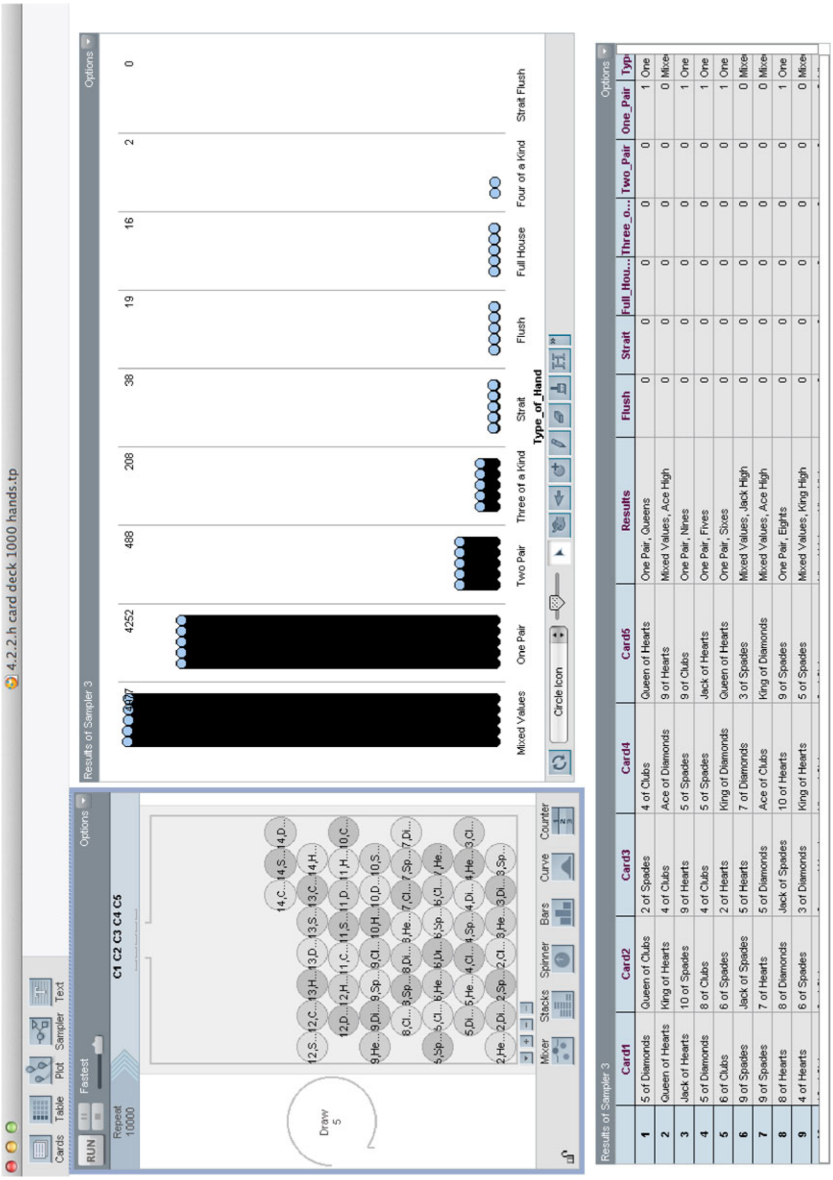


Fig. 10 1000 five-card hands of cards drawn from a fair deck of cards

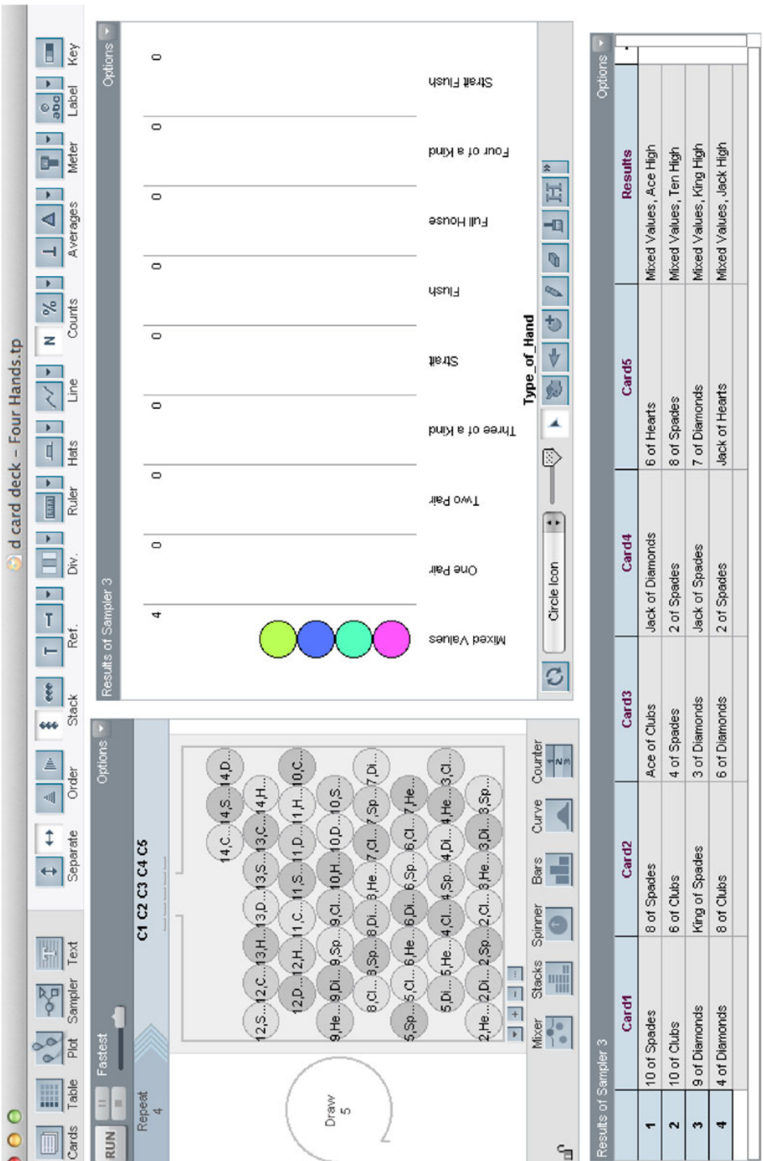


Fig. 11 Four hands of cards drawn from a fair deck

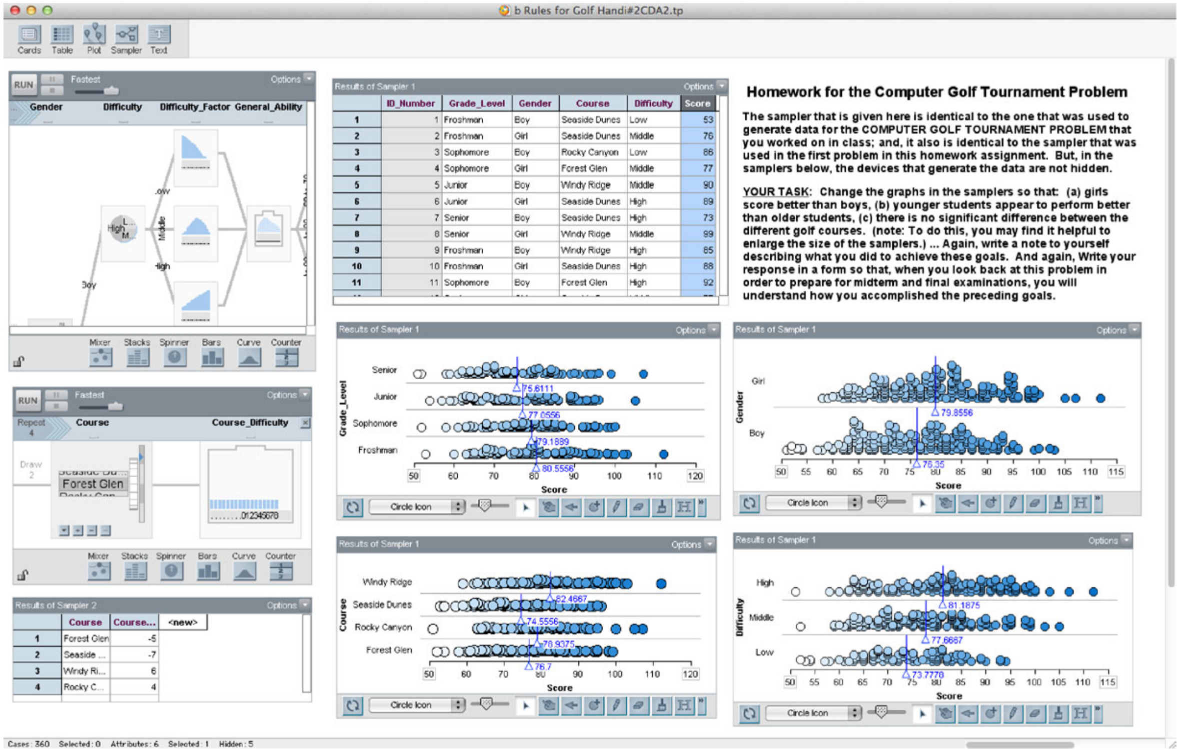
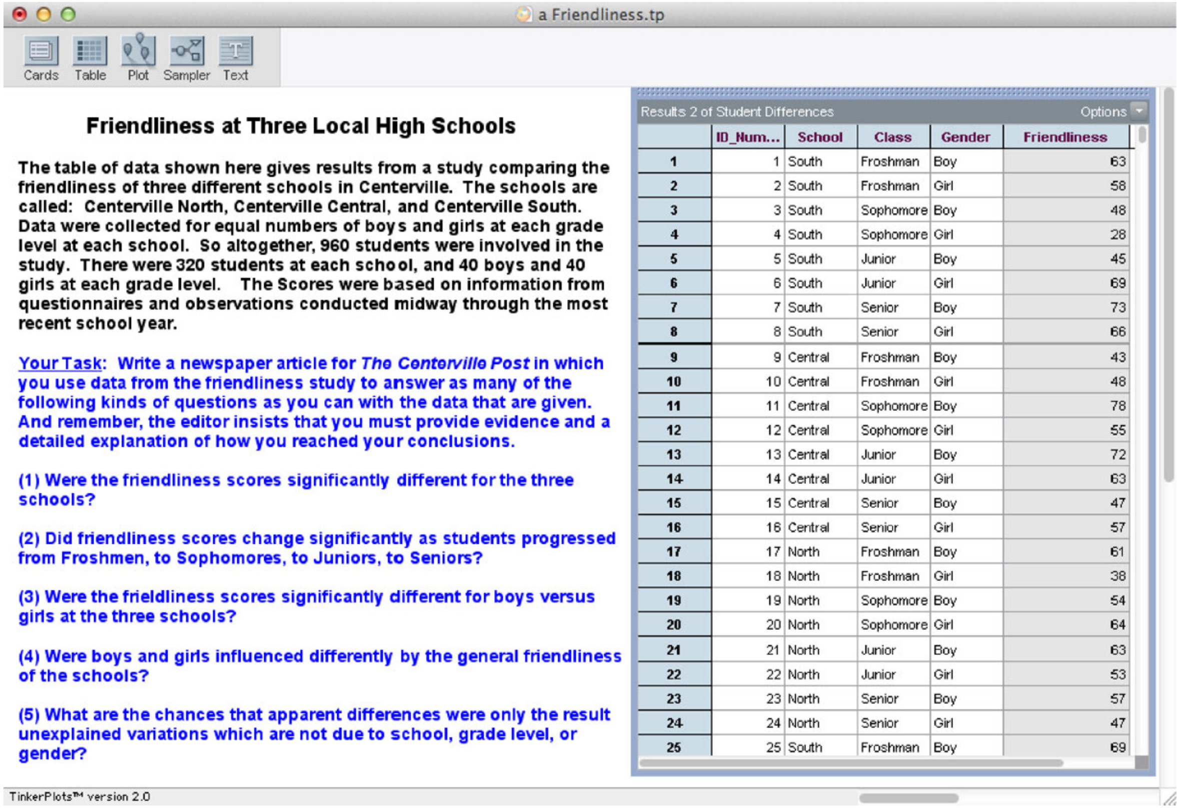


Fig. 12 A computer golf problem



a Friendliness.tp

Cards Table Plot Sampler Text

Friendliness at Three Local High Schools

The table of data shown here gives results from a study comparing the friendliness of three different schools in Centerville. The schools are called: Centerville North, Centerville Central, and Centerville South. Data were collected for equal numbers of boys and girls at each grade level at each school. So altogether, 960 students were involved in the study. There were 320 students at each school, and 40 boys and 40 girls at each grade level. The Scores were based on information from questionnaires and observations conducted midway through the most recent school year.

Your Task: Write a newspaper article for *The Centerville Post* in which you use data from the friendliness study to answer as many of the following kinds of questions as you can with the data that are given. And remember, the editor insists that you must provide evidence and a detailed explanation of how you reached your conclusions.

(1) Were the friendliness scores significantly different for the three schools?

(2) Did friendliness scores change significantly as students progressed from Froshmen, to Sophomores, to Juniors, to Seniors?

(3) Were the frieldliness scores significantly different for boys versus girls at the three schools?

(4) Were boys and girls influenced differently by the general friendliness of the schools?

(5) What are the chances that apparent differences were only the result unexplained variations which are not due to school, grade level, or gender?

Results 2 of Student Differences Options

	ID_Num...	School	Class	Gender	Friendliness
1	1	South	Froshman	Boy	63
2	2	South	Froshman	Girl	58
3	3	South	Sophomore	Boy	48
4	4	South	Sophomore	Girl	28
5	5	South	Junior	Boy	45
6	6	South	Junior	Girl	69
7	7	South	Senior	Boy	73
8	8	South	Senior	Girl	66
9	9	Central	Froshman	Boy	43
10	10	Central	Froshman	Girl	48
11	11	Central	Sophomore	Boy	78
12	12	Central	Sophomore	Girl	55
13	13	Central	Junior	Boy	72
14	14	Central	Junior	Girl	63
15	15	Central	Senior	Boy	47
16	16	Central	Senior	Girl	57
17	17	North	Froshman	Boy	61
18	18	North	Froshman	Girl	38
19	19	North	Sophomore	Boy	54
20	20	North	Sophomore	Girl	64
21	21	North	Junior	Boy	63
22	22	North	Junior	Girl	53
23	23	North	Senior	Boy	57
24	24	North	Senior	Girl	47
25	25	South	Froshman	Boy	69

TinkerPlots™ version 2.0

Fig. 13 A problem comparing the friendliness of three high schools

Why should larger learning environments be needed? And, in fact, why should learning environments be investigated at all if the goal is to study something more than how well students can follow thinking trajectories laid out by authors and teachers? ... One answer to this second question is that, especially when attention shifts toward model-development, the nature of what develops is influenced by both (a) how the authors structured the learning activities, and (b) how students make sense of the situations they are given. So, to understand students' knowledge development in the preceding kinds of situations, it is important to have equally clear descriptions of both of these factors. And, this is why many leading mathematics education researchers have adopted "teaching experiment methodologies" of the type emphasized by a number of authors in the Handbook of Research Design edited by Kelly and Lesh (2000) or "design research methodologies" emphasized in the Handbook of Design Research edited by Kelly et al. (2009). ... One answer to the first questions is that a large share of the meaning of any given "big idea" comes from relationships to other "big ideas" associated facts and skills. In fact, in a given course, one important issue has to do with choices concerning what should be treated as the most important *small number of "big ideas"* around which other concepts, abilities, facts, and skills should be organized. And, in general, it seems reasonable to expect that there is no single "correct" answer to this question. A variety of options tend to exist—especially in courses about probability and statistics—because of changes that have occurred in both the theoretical foundations of the subject, and in the real life situations in which such thinking is needed. Similarly, it also seems reasonable to expect that there is no single "correct" path through whatever conceptual landscape is chosen.

In general, especially when attention shifts toward deeper and higher-order understandings of "big ideas", the state of students' knowledge appears to resemble an ecosystem filled with interacting, partly undifferentiated, conceptual organisms—more than it resembles a collection of increasingly complex computer programs (Lesh and Doerr 2011). Therefore, like the evolutionary factors that govern the development of other types of ecosystems, development requires diversity, selection, communication, and accumulation. And, the kind of courses that embody such factors themselves governed by these same evolutionary factors.

The lack of cumulativeness appears to be the main shortcoming related to the development of productive learning environments (Lesh and Sriraman 2005, 2010). And, curriculum developers need to design for success—not just test for it—similar to the ways that other types of complex systems are designed (for example) in fields like aeronautical engineering. So, for the purposes of both research and curriculum development, the kind of learning environments that are most needed should be modular (so that different specialists can contribute), reconfigurable (for a variety of circumstances), adaptable (for a variety of purposes). There are strong reasons to reject the notion that there exists one single non-adaptable learning environment that fits the needs of all students, all teachers, and all learning communities. And, there are equally strong reasons to assume that the kind of learning environments that are most needed by both researchers and practitioners will need to be continually evolving and easily adaptable to fit the needs of a variety of people, situations, and purposes.

References

Borromeo-Ferri, R., & Lesh, R. (2011). Should interpretation systems be considered to be models if they only function implicitly? In G. Kaiser, W. Blum, & K. Maass (Eds.), *Teaching mathematical modeling & applications*. Berlin: Springer.

Common Core State Standards Initiative (2010). Common Core State Standards for mathematics. Retrieved from http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf.

Doerr, H., & Lesh, R. (2010). Model-eliciting activities: do they work in the education of students, teams, or teachers? In G. Kaiser, W. Blum, & K. Maass (Eds.), *Teaching mathematical modeling & applications*. Berlin: Springer.

Hamilton, E., Lesh, R., Lester, F., & Yoon, C. (2007). The use of reflection tools in building personal models of problem solving. In R. Lesh, E. Hamilton, & J. Kaput (Eds.), *Foundations for the future in mathematics education* (pp. 349–366). Mahwah: Erlbaum.

Kelly, A., & Lesh, R. (Eds.) (2000). *The handbook of research design in mathematics and science education*. Hillsdale: Erlbaum.

Kelly, A. E., Lesh, R. A., & Baek, J. (2009). *Handbook of design research methods in education*. New York: Routledge.

Lesh, R. (2002). Research design in mathematics education: focusing on design experiments. In L. English (Ed.), *International handbook of research design in mathematics education*. Hillsdale: Erlbaum.

Lesh, R. (2003a). Multi-tier design research. In L. English, M. Bartolini Bussi, G. Jones, R. Lesh, B. Sriraman, & D. Tirosh (Eds.), *Handbook of international research in mathematics education*. New York: Routledge.

Lesh, R. (2003b). *Monograph for international journal for mathematical thinking & learning. Models & modeling in mathematics education*. Hillsdale: Erlbaum.

Lesh, R., & Caylor, E. (2007). Modeling as application vs modeling as a way to create mathematics. *International Journal of Computers for Mathematical Learning*, *12*, 173–194.

Lesh, R., & Doerr, H. (2003). *Beyond constructivist: a models & modeling perspective on mathematics teaching, learning, and problems solving*. Hillsdale: Erlbaum.

Lesh, R., & Doerr, H. (2011). Alternatives to trajectories and pathways to describe development in modeling and problem solving. In R. Borromeo-Ferri (Ed.), *Teaching mathematical modeling & applications*. Berlin: Springer.

Lesh, R., & Lamon, S. (1992). *Assessing authentic mathematical performance*. Washington: American Association for the Advancement of Sciences Press.

Lesh, R., & Sriraman, B. (2005). Mathematics education as a design science. *International Reviews on Mathematical Education*, *37*(6), 490–505.

Lesh, R., & Sriraman, B. (2010). Mathematics education as a design science. In B. Sriraman & L. English (Eds.) *Advances in mathematics education: Vol. 1. Theories of mathematics education: seeking new frontiers* Berlin: Springer.

Lesh, R., & Yoon, C. (2004). Evolving communities of mind—in which development involves several interacting and simultaneously developing strands. *Mathematical Thinking and Learning*, *6*(2), 205–226.

Lesh, R., & Zawojewski, J. (2007). Problem solving and modeling. In F. K. Lester Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 763–804). Charlotte: Information Age Publishing.

Lesh, R., Hamilton, E., & Kaput, J. (Eds.) (2007). *Models & modeling as foundations for the future in mathematics education*. Hillsdale: Erlbaum.

Lesh, R., Carmona, G., & Moore, T. (2010a). Six sigma learning gains and long term retention of understanding and attitudes related to models & modeling. *Mediterranean Journal for Research in Mathematics Education*, *9*(1), 19–54.

Lesh, R., Haines, C., Galbraith, P., & Hurford, A. (2010b). *Modeling students' mathematical modeling competencies*. Berlin: Springer.

Piaget, J. (1970). *Genetic epistemology*. New York: Norton.

Sriraman, B., & Lesh, R. (2007). Leaders in mathematical thinking & learning—a conversation with Zoltan P. Dienes. *Mathematical Thinking and Learning: An International Journal*, *9*(1), 59–75.

Vygotsky, L. S. (1962). *Thought and language*. Cambridge: Massachusetts Institute of Technology Press.

Zawojewski, J., Diefes-Dux, H., & Bowman, K. (Eds.) (2009). *Models and modeling in engineering education: designing experiences for all students*. Rotterdam: Sense Publications.

Perspective III: Stochastics

Preface to Perspective III: Stochastics

Egan J. Chernoff and Gale L. Russell

A university professor, teaching an undergraduate mathematics course involving probability and statistics, gives his students the following task (before he leaves the room!): Half of the class is asked to flip a coin 100 times and record the resulting sequence of heads and tails on the blackboard; the other half of the class is tasked with "cooking up" a sequence of 100 heads and tails (without flipping a coin) and also recording the results on the blackboard. Upon his return, to everyone's surprise, the professor immediately (and correctly) identifies which sequence was derived from flipping the coin and which was made-up. How is this possible?

The term "stochastics" has different meanings for different individuals. Despite a multitude of meanings, a commonly accepted "definition" is simply: probability and statistics. However, we contend the above story—for many years considered urban legend, but now actualized in introductory statistics courses around the world (e.g., Deborah Nolan and others)—best embodies the notion of stochastics in mathematics education. From this contention, by explicitly recognizing the notion of randomness within stochastics, we (re)define (for the purposes of this preface) stochastics, simply, as randomness and probability and statistics.

A result of the different mathematical notions of randomness (and akin to the term "stochastics"), the term "randomness" has different meanings for different individuals. This terminological confusion is compounded in the field of mathematics education because research focuses less on the different (mathematical) notions of randomness and more on perceptions of randomness or "subjective perceptions of randomness" (Batanero, Arteaga, Serrano & Ruiz, this volume). Drawing heavily on the research of psychologists, researchers in mathematics education have consistently investigated perceptions of randomness. This volume reveals certain interesting developments to this domain of research (e.g., Jolfaee, Zazkis & Sinclair). These developments are in line with the forward looking nature of the *Advances in Mathematics Education* series and will inevitably inform research investigating the other two (inextricably linked) notions of stochastics: probability and statistics.

E.J. Chernoff (✉) · G.L. Russell
University of Saskatchewan, Saskatoon, Saskatchewan, Canada
e-mail: egan.chernoff@usask.ca

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_18,

As statistics continues to move onto the world's stage (e.g., Big Data, Nate Silver), so too it is quickly becoming a darling of mathematics education (with dedicated journals and conferences). Further, certain developments associated with statistics on the world's stage are being paralleled in statistics research in the field of mathematics education. For example, statistical literacy, data modeling, technology and simulation, to name but a few, are topics of recent focus in statistics and are also investigated throughout this perspective.

Well documented, the field of statistics was (and many will argue still is) segregated into two camps: Bayesians and Frequentists. Worthy of note, the field of mathematics education is not immune from this division—elements of which are witnessed in this perspective.

Currently, there exists a hierarchical relationship in the volume of stochastics research in the field of mathematics education. Specifically, research investigating statistics is most prevalent, followed by research investigating probability and, lastly, followed by research investigating (perceptions of) randomness. The chapters found within this perspective may lead to, at some point in the future, a shift in the current hierarchy that exists.

After all, looking back, what captures our attention in the original story is not the sequence derived from 100 flips of a fair coin or, for that matter, the "cooked up" sequence of heads and tails; rather, it is the role of the individual in the juxtaposition of the two sequences, which drives our research into (subjective) probability, (perceptions) of randomness and (informal) inference.

Prospective Primary School Teachers' Perception of Randomness

Carmen Batanero, Pedro Arteaga, Luis Serrano, and Blanca Ruiz

Abstract Subjective perception of randomness has been researched by psychologists and mathematics educators, using a variety of tasks, resulting in a number of different descriptions for the biases that characterize people's performances. Analysing prospective teachers' possible biases concerning randomness is highly relevant as new mathematics curricula for compulsory teaching levels are being proposed that incorporate increased study of random phenomena. In this chapter, we present results of assessing perception of randomness in a sample of 208 prospective primary school teachers in Spain. We first compare three pairs of random variables deduced from a classical task in perception of randomness and deduce the mathematical properties these prospective teachers assign to sequences of random experiments. Then, the written reports, where prospective teachers analyse the same variables and explicitly conclude about their own intuitions are also studied. Results show a good perception of the expected value and poor conception of both independence and variation as well as some views of randomness that parallel some naïve conceptions on randomness held at different historic periods.

1 Introduction

Different reasons to teach probability have been highlighted over the past years (e.g. by Gal 2005; Franklin et al. 2005; Jones 2005; and Borovcnik 2011): the role of probability reasoning in decision making, the instrumental need of probability in other disciplines, and the relevance of stochastic knowledge in many professions. Moreover, students meet randomness not only in the mathematics classroom, but also in social activities (such as games or sports), and in meteorological, biological, economic, and political settings. Consequently, some of them may build incorrect conceptions of randomness in absence of adequate instruction (Borovcnik 2012).

C. Batanero (✉) · P. Arteaga · L. Serrano
University of Granada, Granada, Spain
e-mail: batanero@ugr.es

B. Ruiz
ITESM, Monterrey, Mexico

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_19,

As a consequence of this pervasive presence of chance in everyday life, probability has been included in recent curricula of primary school education that place more emphasis on the study of randomness and probability by very young children. For example, in the Spanish curriculum for compulsory primary education (MEC 2006), we find the following contents in the first cycle (6–7 year-olds): "*Random nature of some experiences. Difference between possible, impossible and that what is possible but not certain*." Reference is made to using the chance language in everyday settings, in order to describe and quantify random situations. In the second cycle (8–9 year olds), the document suggests that children should evaluate the results of random experiences, and understand that there are more and less probable events, and that it is impossible to predict a specific result. In the last cycle (10–11 year olds), children are encouraged to recognise random phenomena in everyday life and estimate the probability for events in simple experiments.

This curriculum is not an exception, since the current tendency even for primary school levels is towards data-orientated teaching of probability, where students are expected to perform experiments or simulations, formulate questions or predictions, collect and analyse data from these experiments, propose and justify conclusions and predictions that are based on data (e.g. Ministério da Educação 1997; NCTM 2000).

Changing the teaching of probability in schools will depend on the extent to which we can adequately prepare the teachers. Although teachers do not need high levels of mathematical knowledge, they do require a profound understanding of the basic mathematics they teach at the school level, including a deep grasp of the interconnections and relationships among different aspects of this knowledge (Ma 1999); for example, teachers need a sound understanding of the different meanings associated to randomness and probability. Unfortunately, several authors (e.g. Franklin and Mewborn 2006; Chadjipadelis et al. 2010; Batanero and Díaz 2012) suggest that many of the current programmes do not yet train teachers adequately for their task to teach statistics and probability. The situation is particularly challenging for primary school teachers, few of whom have had suitable training in statistics, or in the related didactical knowledge (Jacobbe 2010) and, consequently, they might share with their students a variety of probabilistic misconceptions (Stohl 2005). Therefore, it is important to assess teachers' probabilistic knowledge and find activities where teachers work with meaningful problems and are confronted to their own misconceptions in the topic (Batanero et al. 2004).

Understanding randomness is the base of understanding probability and conceptions of randomness are at the heart of people's probabilistic and statistical reasoning (Lecoutre et al. 2006); however, epistemological analysis of randomness, as well as psychological research, has shown that there is no adequate perception of randomness in children or even in adults. There is an apparent contradiction in people's understanding of random processes and sequences, which is related to the psychological problems associated with the concept, namely that randomness implies that "anything possible might occur", but that subjectively many people believe, however, that only the outcomes without visible patterns are "permissible" examples of randomness (Hawkins et al. 1991).

Despite the relevance of the topic in probability and statistics, little attention has been paid to prospective teachers' conceptions on randomness. To address this omission, in this chapter, we analyse research that was aimed at assessing prospective primary school teachers' perception of randomness using two different tools: (a) we first analyse some statistical variables deduced from a classical experiment related to perception of randomness that was carried out by the teachers; (b) we secondly analyse the written reports produced by the teachers, which were part of an activity directed to confront them with their own misconceptions of randomness.

In the next sections, we first analyse some different historical interpretation of randomness that can be parallel to some conceptions shown by prospective teachers in our research. We secondly analyse previous research on subjective perceptions of randomness in children and adults, and the scarce research dealing with teachers. Then we present the method, results and conclusions of our study. Finally, some implications for teachers' education are provided.

2 Randomness: Emergence and Progressive Formalization

In spite of being a basic idea in probability, randomness is not an easy concept. The term resists easy or precise definition, its emergence was slow and it has received various interpretations at different periods in history (Zabell 1992; Bennet 1998; Liu and Thompson 2002; Batanero et al. 2005). Some of these interpretations are relevant to this research and may help understanding prospective teachers' difficulties in the theme.

2.1 Randomness and Causality

Chance mechanisms such as cubic dice or astragali have been used since antiquity to make decisions or predict the future. However, a scientific idea of randomness was absent in the first exploratory historical phase, which extended according to Bennet (1998), from antiquity until the beginning of the Middle Ages when randomness was related to causality and conceived as the opposite of something that had some known causes. According to Liu and Thompson (2002), conceptions of randomness and determinism ranged along an epistemological spectrum, where, on the one extremum, random phenomena would not have an objective existence but would reflect human ignorance. This was the view, for example, of Aristotle who considered that chance results from the unexpected coincidence of two or more series of events, independent of each other and due to so many different factors that the eventual result is pure chance (Batanero et al. 2005). It was also common in European Enlightenment where there was a common belief in universal determinism, as expressed, for example, by Laplace: "We ought then to regard the present state of the universe as the effect of its anterior state and as the cause of the one which is to follow" (Laplace 1995, p. vi). From this viewpoint, chance is seen as only the expression of our ignorance.

The other end of the spectrum consists in accepting the existence of "irreducible chance" and therefore that randomness is an inherent feature of nature (Liu and Thompson 2002). As stated by Poincaré (1936), for the laws of Brownian motion, the regularity of macroscopic phenomena can be translated to deterministic laws, even when these phenomena are primarily random at the microscopic level. Moreover, ignorance of the laws governing certain natural phenomena does not necessarily involve a chance interpretation because certain phenomena with unknown laws (such as death) are considered to be deterministic. Finally, among the phenomena for which the laws are unknown, Poincaré discriminated between random phenomena, for which probability calculus would give us some information, and those non-random phenomena, for which there is no possibility of prediction until we discover their laws. As Ayer (1974) stated, a phenomenon is only considered to be random if it behaves in accordance with probability calculus, and this definition will still hold even when we have found the rules for the phenomenon.

2.2 *Randomness and Probability*

With the pioneer developments of probability theory, randomness was related to equiprobability (for example, in the Liber of Ludo Aleae by Cardano) because this development was closely linked to games of chance where the number of possibilities is finite and the principle of equal probabilities for the elementary events of the sample space in a simple experiment is reasonable.

Nowadays, we sometimes find randomness explained in terms of probability, although such an explanation would depend on the underlying understanding of probability. If we adopt a Laplacian view of probability, we would consider that an object is chosen at random out of a given class (sample space), if the conditions in this selection allow us to give the same probability for any other member of this class (Lahanier-Reuter 1999). This definition of randomness may be valid for random games based on dice, coins, etc., but Kyburg (1974) suggested that it imposes excessive restriction to randomness and it would be difficult to find applications of the same. For example, it only can be applied to finite sample spaces; if the sample space is infinite, then the probability associated to each event is always null, and therefore still identical, even when the selection method is biased. Furthermore, this interpretation precludes any consideration of randomness applied to elementary events that are not equiprobable.

When we transfer the applications of probability to the physical or natural world, for example, studying the blood type of a new-born baby or any other hereditary characteristic, we cannot rely on the equiprobability principle. In this new situation, we may consider an object as a random member of a class if we can select it using a method providing a given 'a priori' relative frequency to each member of this class in the long run. Thus, we use the frequentist view of probability, which is most appropriate when we have data from enough cases. However, we are left with the theoretical problem of deciding how many experiments are necessary to be considered in order to be sure that we have sufficiently proven the random nature of the object (Batanero et al. 2005).

Within either of these two frameworks, randomness is an 'objective' property assigned to the event or element of a class. Kyburg (1974) criticizes this view and proposes a subjective interpretation of randomness composed of the following four terms:

- The object that is supposed to be a random member of a class;
- The set of which the object is a random member (population or collective);
- The property with respect to which the object is a random member of the given class;
- The knowledge of the person giving the judgement of randomness.

Whether an object is considered to be a random member of a class or not depends, under this interpretation, on our knowledge. Consequently, this view is coherent with the subjective conception of probability and is adequate when we have some information affecting our judgement about the randomness of an event (Fine 1973).

2.3 Formalization of Randomness

By the end of the nineteenth century, theoretical developments of statistical inference and publication of tables of pseudorandom numbers produced concern about how to ensure the 'quality' of those numbers. According to Zabell (1992), an important development was the distinction between a random process and a random sequence of outcomes: Although randomness is a property of a process, rather than of the outcomes of that process, it is only by observing outcomes that we can judge whether the process is random or not (Johston-Wilder and Pratt 2007). The possibility of obtaining pseudorandom digits with deterministic algorithms and related debates led to the formalization of the concept of randomness (Fine 1973).

Von Mises (1952) based his study of this topic on the intuitive idea that a sequence is considered to be random if we are convinced of the impossibility of finding a method that lets us win in a game of chance where winning depends on forecasting that sequence. This definition of randomness is the basis for statistical tests that are used for checking random number tables before presenting them to the scientific community. However, since there is the possibility of error in all statistical tests, we can never be totally certain that a given sequence, in spite of having passed all the tests, does not have some unnoticed pattern within it. Thus, we cannot be absolutely sure about the randomness of a particular finite sequence. We only take a decision about its randomness with reference to the outcomes of test techniques and instruments. This explains why a computer-generated random sequence (which is not random in an absolute sense) can still be random in a relative sense (Harten and Steinbring 1983).

Another attempt to define the randomness of a sequence was based on its *computational complexity*. Kolmogorov's interpretation of randomness reflected the difficulty of describing it (or storing it in a computer) using a code that allows us to reconstruct it afterwards (Zabell 1992). In this approach, a sequence would be random if it cannot be codified in a more parsimonious way, and the absence of patterns is its essential characteristic. The minimum number of signs necessary to code a particular sequence provides a scale for measuring its complexity, so this definition allows for a hierarchy in the degrees of randomness for different sequences. It is important to remark that in both von Mises' and Kolmogorov's approaches perfect randomness would only apply to sequences of infinite outcomes and therefore, randomness would only be a theoretical concept (Fine 1973).

3 Perception of Randomness

3.1 Adult's Perception of Randomness

There has been a considerable amount of research into adults' subjective perception of randomness (e.g. Wagenaar 1972; Falk 1981; Bar-Hillel and Wagenaar 1991; Engel and Sedlmeier 2005). Psychologists have used a variety of stimulus tasks which were classified in a review by Falk and Konold (1997) into two main types: (a) In *generation tasks*, subjects generate random sequences under standard instructions to simulate a series of outcomes from a typical random process, such as tossing a coin; (b) In *recognition tasks*, which were termed as *comparative likelihood task* by Chernoff (2009a), people are asked to select the most random of several sequences of results that might have been produced by a random device or to decide whether some given sequences were produced by a random mechanism. Similar types of research have also been performed using two-dimensional random distributions, which essentially consist of random distributions of points on a square grid.

One main conclusion of all these studies is that humans are not good at either producing or perceiving randomness (Falk 1981; Falk and Konold 1997; Nickerson 2002), and numerous examples demonstrate that adults have severe difficulties when appropriately dealing with aspects of randomness, so that systematic biases have consistently been found. One such bias is known as the *gambler's fallacy*, or belief that the probability of an event is decreased when the event has occurred recently, even though the probability is objectively known to be independent across trials (Tversky and Kahneman 1982). Related to this is the tendency of people in sequence generation tasks to include too many alternations of different results (such as heads and tails in the flipping of a coin) in comparison to what would theoretically be expected in a random process. Similarly, in perception tasks people tend to reject sequences with long runs of the same result (such as a long sequence of heads) and consider sequences with an excess of alternation of different results to be random (Falk 1981; Falk and Konold 1997). Comparable results are found in people's performance with two-dimensional tasks in which clusters of points seem to prevent a distribution from being perceived as random. These results support the findings by Tversky and Kahneman (1982) that suggest people do not follow the principles of probability theory in judging the likelihood of uncertain events and apply several heuristics, such as local representativeness, where subjects evaluate the probability of an event by the degree of likeness to the properties of its parent population or the degree it reflects the features of the process by which it is generated. Other authors (e.g. Falk 1981; and Falk and Konold 1997) believe that individual consistency in people's performance with diverse tasks suggests underlying misconceptions about randomness.

3.2 Children's Perception of Randomness

From a didactic point of view, a crucial question is whether these biases and misconceptions are spontaneously acquired or whether they are a consequence of poor instruction in probability. Below, we outline a number of key research studies looking at children's and adolescents' conceptions of randomness, and their performance when faced with tasks requiring the generation or recognition of sequences of random results.

According to Piaget and Inhelder (1951), chance is due to the interference of a series of independent causes, and the 'non-presence' of all the possible outcomes when there are only a few repetitions of an experiment. Each isolated case is indeterminate or unpredictable, but the set of possibilities may be found using combinatorial reasoning, thereby making the outcome predictable. The authors' notion of probability is based on the ratio between the number of possible ways for a particular case to occur and the number of all possible outcomes. This theory would suggest that chance and probability cannot be totally understood until combinatorial and proportional reasoning is developed which, for Piaget and Inhelder, does not happen until a child reaches the formal operations stage (12–14 years).

Piaget and Inhelder (1951) investigated children's understanding of patterns in two-dimensional random distributions. They designed a piece of apparatus to simulate rain drops falling on paving stones. The desire for regularity appeared to dominate the young children's predictions. When they were asked where the following rain drop would fall, children at stage 1 (6 to 9 years) allocated the rain drops in approximately equal numbers on each pavement square, thereby producing a uniform distribution. With older children, proportional reasoning begins to develop, and Piaget and Inhelder reported that such children tolerate more irregularity in the distribution. The authors believed that children understood the law of the large numbers, which explains the global regularity and the particular variability of each experiment simultaneously.

Fischbein and Gazit (1984) and Fischbein et al. (1991) have also documented children's difficulties in differentiating random and deterministic aspects, and their beliefs in the possibility of controlling random experiments. In contrast to the Piagetian view, these authors have suggested that even very young children display important intuitions and precursor concepts of randomness and consequently argue that it is not didactically sound to delay exploiting and building on these subjective intuitions until the formal operations stage is reached.

Moreover, Green's (1983) findings, also contradicted Piaget and Inhelder' theory. His investigations with 2930 children aged 11–16, using paper and pencil versions of Piagetian tasks, showed that the percentage of children recognising random or semi-random distributions actually decreased with age. In a second study with 1600 pupils aged 7 to 11 and 225 pupils aged 13 to 14 (Green 1989, 1991), Green gave the children generation and recognition tasks related to a random sequence of heads and tails representing the results of flipping a fair coin. The study demonstrated that children were able to describe what was meant by equiprobable. However, they did not appear to understand the independence of the trials and tended to produce series in which runs of the same result were too short compared to those that we would expect in a random process. In both studies, children based their decisions on the following properties of the sequences: results pattern, number of runs of the same result, frequencies of results, and unpredictability of random events. However, these properties were not always correctly associated to randomness or determinism.

Toohey (1995) repeated some of Green's studies with 75 11–15 year-old students and concluded that some children have only a *local perspective* of randomness, while that of other children is entirely global. The local perspective of randomness emphasizes the spatial arrangement of the outcomes within each square, while the *global perspective* concentrates on the frequency distribution of outcomes.

Batanero and Serrano (1999) analysed the written responses of 277 secondary school students in two different age groups (14 and 17 year-olds) to some test items taken from Green (1989, 1991) concerning the perception of randomness in sequences and two-dimensional distributions. The authors also asked the students to justify their answers. Batanero and Serrano's results suggest that students' subjective meaning of randomness could parallel some interpretations that randomness has received throughout history. For example, when a student associated the lack of pattern to randomness, Batanero and Serrano suggested this view was consistent with the *complexity approach* to randomness described before. Other students showed a conception compatible with, the *classical*, *frequentist or subjective approach* to probability and randomness.

More recently, researchers have used computers to simulate random processes in order to discover children's understanding of randomness. Pratt (2000) analysed 10–11 year-old children's ideas of randomness as they worked in a computer environment and suggested these children showed the local and global perspective of randomness described by Toohey (1995). While the local perspective children's attention is mainly paid to the uncertainty of the next outcome and the ephemeral patterns in short sequences, in the global view the children were aware of the long term predictability of either the empirical distribution of outcomes (frequency of observed outcomes) or the theoretical distribution (expressing beliefs about the behaviour of the random generator, such as equally likelihood of different outcomes) (see also Johston-Wilder and Pratt 2007).

3.3 Teachers' Perception of Randomness

In this section, we summarise the scarce research focussed on teachers' perception of randomness which suggests the need to develop specific training where teachers can increase their probabilistic knowledge for teaching.

Azcárate et al. (1998) analysed the responses of 57 primary-school teachers to Konold et al. (1991) questionnaire in order to analyse these teachers' conception of randomness. They also asked participants to list examples or random and non-random phenomena and to describe the features they assigned to random phenomena. In general, participants showed a partial conception of randomness, which reflected, in most cases, causal argumentations and poor perception of random processes in everyday settings (beyond games of chance). Many participants considered a phenomena to be deterministic if they could identify some causes that could influence the phenomena apart pure chance (e.g. in meteorology). Other criteria to judge randomness included multiple possibilities or unpredictability of results.

The most relevant study with teachers is that by Chernoff (2009a, 2009b). After a pilot study with 56 prospective teachers, Chernoff (2009a) analysed the responses and justifications given by 239 prospective mathematics teachers (163 elementary school teachers and 76 secondary school teachers) to a questionnaire consisting in comparative likelihood tasks. The questionnaire included several sequences of 5 trials of flipping a fair coin, in which the author fixed the ratio of heads to tails and varied the arrangement of outcomes. In order to show that responses that were assumed as incorrect in previous research could be derived from participants' subjective probabilistic thinking, Chernoff (2009b) analysed the justifications of 19 prospective teachers that apparently had incorrect perception of randomness. The result of his analysis suggested that these prospective teachers may be reasoning from three different interpretations of the sample space: (a) taking into account the switches from head to tail; (b) considering the longest run; and (c) considering the switches and longest run together. Consequently, their reasoning as regards randomness and their judgement of whether a sequence was random or not could be consistent with these views of the sample space; therefore their apparent incorrect responses were not due to lack of probabilistic reasoning, but to use of personal subjective probabilities.

In summary, research carried out with prospective teachers is scarce and suggest a poor perception of randomness as well as use of subjective probabilities. Below we summarise our own research in which we explore the teachers' capability to judge their own intuitions when analysing the data collected by themselves in a generation task and the possibility that part of the teachers show some naïve conceptions on randomness that parallel those held at different historic periods.

4 Method

Participants in our study were 208 prospective primary school teachers in the Faculty of Education, University of Granada, Spain; in total 6 different groups (35–40 prospective teachers by group) took part in this research. All of these prospective teachers (in their second year at the university) were following the same mathematics education course, using the same materials and doing the same practical activities. They all had followed a mathematics course, which included descriptive statistics and elementary probability the previous year.

The data were collected as a part of a formative activity, which is discussed in depth in Godino et al. (2008) and consisted of two sessions (90 minutes long each). The two main goals of the formative activity were: (a) assessing prospective teachers' conceptions of randomness; (b) confronting prospective teachers with their possible misconceptions on this concept.

In the first session (90 minutes long), the prospective teachers were given the statistical project "Check your intuitions about chance" (Godino et al. 2008) in which they were encouraged to carry out an experiment to decide whether the group had good intuition of randomness or not. The experiment consisted of trying to write down apparent random results of flipping a coin 20 times (without really throwing the coin, just inventing the results) in such a way that other people would think the coin was flipped at random (simulated sequence). This is a classical generation task which was similar to that used in Engel and Sedlmeier's (2005) research.

Participants recorded the simulated sequences on a recording sheet. Afterwards they were asked to flip a fair coin 20 times and write the results on the same recording sheet (real sequence). At the end of the session, in order to confront these future teachers with their misconceptions, participants were given the data collected in their classroom. These data contained six statistical variables: number of heads, number of runs and length of the longest run for each of real and simulated sequences from each student. Results of the experiments are presented in Figs. 1 through 3. Sample sizes for the data analysed by the prospective teachers in each group were smaller (30–40 experiments per group), although the shape of the distribution and summaries for each variable were very close to those presented in Figs. 1 through 3.

Teachers were asked to compare the variables collected from the real and simulated sequences, finish the analysis at home and write a report with a complete discussion of the project, including all the statistical graphs and procedures they used and their conclusions regarding the group's intuitions about randomness. Participants were given freedom to build other graphs or summaries in order to complete their reports. In the second session, the reports were collected and the different solutions to the project given by the prospective teachers were collectively discussed in the classroom. In addition, a didactical analysis was carried out in order to reflect on the statistical knowledge needed to complete the project and on the pedagogical content knowledge involved in teaching statistics at a primary school through project work.

5 Results and Discussion

In order to assess prospective teachers' conceptions of randomness, we first analysed the number of heads, number of runs and longest run in each of the simulated

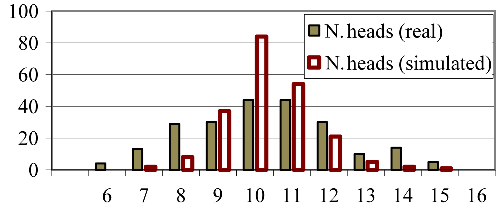


Fig. 1 Distribution for number of heads

and real sequences in the data collected by the prospective teachers in their experiments. In the following, the data collected by the six groups taking part in the study ($n = 208$) will be analysed together, although similar results were found in each of the six subsamples.

5.1 Perception of the Binomial Distribution

The theoretical distribution for the number of heads in 20 trials can be modelled by the binomial distribution $B(n, p)$, where $n = 20$ and $p = 0.5$. The expectation and variance for this distribution are $\mu = np = 10$ and $Var = npq = 5$. The empirical distributions for the number of heads in the real and simulated sequences in the experiments carried out by the participants in the study are presented in Fig. 1 and the summary statistics in Table 1.

We can observe that participants produced "heads" and "tails" in about equal numbers, and the majority of the students were close to the theoretically correct expected value of 10. There was no significant preference for "heads" over "tails" or vice versa, in agreement with previous research (e.g. Wagenaar 1972; Green 1991; Falk and Konold 1997; Engel and Sedlmeier 2005).

Results show a good perception of the expected number, median and mode of the binomial distribution (number of heads in 20 flips of a coin) as it is shown in the mode, median and average number of heads in the simulated sequences, which are close to the theoretical value $np = 10$ and in the non-significant difference in the t-test of difference of averages between real and simulated sequences ($t = -1.00$; $p = 0.31$). The standard deviation in the simulated sequences was, however, almost

Table 1 Summary statistics for number of heads, longest run and number of runs

	Number of heads		Longest run		Number of runs	
	Real	Simulated	Real	Simulated	Real	Simulated
Mean	10.45	10.29	4.35	3.32	10.10	10.78
Mode	10; 11	10	4	3	10; 11	12
Median	10	10	4	3	10	12
Std. deviation	2.05	1.22	1.6	1.12	2.9	2.8
Range	11	8	10	11	14	13

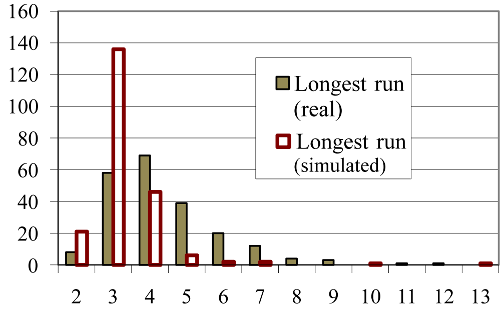


Fig. 2 Distribution and summary statistics for the longest run

half the theoretical value and the differences were statistically significant in the F-test ($F = 2.83$, $p = 0.001$).

5.2 Perception of Independence

Even if the concept of independence is very easy to define mathematically by the property that a joint probability of two events is the product of the two single event probabilities, its application is difficult for many people. Perception of independence was poor in our sample, as prospective teachers produced on average shorter runs and higher number of runs than expected in a random process (according to Engel and Sedlmeier 2005, the expected number of runs in a coin flipping sequence of length 20 is 10.5, and the expected longest run 4.33). This is consistent with previous research, in which, a majority of participants of different ages—examined under various methods—identify randomness with an excess of alternations between different outcomes (e.g. Wagenaar 1972; Green 1991; Engel and Sedlmeier 2005) and whereas the expected probability of alternation in a random binary sequence is 0.5, people's average subjective preference is a probability of 0.6 or 0.7 in generation tasks (Falk and Konold 1997). Moreover, according Tversky and Kahneman's (1982) representativeness heuristic, people judge that a sequence is random if it is representative of its parent population; therefore, the sequences comprise about equal numbers of heads and tails and display an irregular order, not only globally, but also locally. This result is visible in Figs. 2 and 3 and in the t-tests of differences of averages that was statistically significant for both variables ($t = -7.76$; $p = 0.001$ for the longest run; $t = 2.48$; $p = 0.01$ for the number of runs).

Some teachers recognised this difference in their reports and consequently they correctly considered this was an indication of an incomplete perception of randomness: "In the simulated sequence, we tend to produce many short runs" (AB). On the contrary, misconceptions of independence, was also observed in some written reports by other participants, who rejected the sequence as random because some runs were longer than what they believed should be expected in a random sequence: "Some students cheated and invented their sequences, since they produced too many successive heads or tails to be random" (EA).

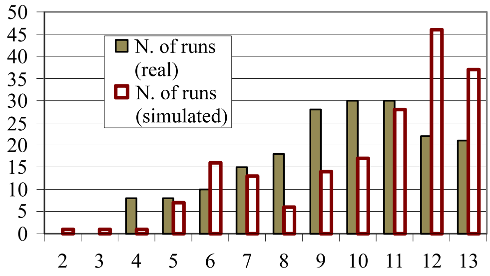


Fig. 3 Distribution and summary statistics for the number of runs

5.3 Perception of Variation

Variability is omnipresent throughout the statistical enquiry cycle and is fundamental to statistical thinking (Wild and Pfannkuch 1999). However, the prospective teachers in our study produced sequences with little variation in the length of runs (a wide majority produced a longest run with only 3 similar outcomes). This is visible in Fig. 2 and in the F-test that was statistically significant ($F = 2.06$; $p = 0.0001$). However, the perception of variation was good as regards the number of runs ($F = 1.07$; $p = 0.6$; not significant). This result is reasonable since participants were not committed, as a group, to reproduce the sampling distribution of the proportions of heads. Consequently, their deviation from the expected variability in the distribution of all the random sequences is a result of in each individual's matching of the expected proportion in his/her sequence produced (Falk et al. 2009). However, in the next sections, we will show that the perception of variation in these prospective teachers was poor since only a few of them made reference to variation in their reports as an important feature of random processes.

Anyway, part of the teachers perceived variation in the data, and even more, justified randomness basing on this variation "There is more variety in the random sequence. This is pretty logical, since these results were obtained by a random experiment that involved chance" (NC); " In the number of heads, there is a difference, ... since when we made up the data (in the simulated sequence) results were more even, but in the real sequence these results are more uneven, since they are due to chance" (IE).

5.4 Teachers' Analyses of Their Own Intuitions

All the above results reproduced those obtained by Green (1991) and Batanero and Serrano (1999) with secondary school students, which is reasonable because the statistics training that Spanish prospective primary teachers receive is reduced to their study of statistics along secondary education.

As we have previously explained, after performing the experiment, participants in the study were given a data sheet with the data recorded in the classroom and were asked to analyse these data and conclude on their own intuitions in a written

Table 2 Frequency (percent) of prospective teachers' conclusions

Conclusion about the intuitions ($n = 200$)		Correct	Incorrect	No conclusion
Perceiving the expected value	Number of heads	32 (16.0)	114 (57.0)	54 (27.0)
Perceiving independence	Expected longest run	6 (3.0)	87 (43.5)	107 (53.5)
	Expected number of runs	9 (4.5)	100 (50.0)	91 (45.5)
Perceiving variation	Number of heads	25 (12.5)	121 (60.5)	54 (27.0)
	Longest run	8 (4.0)	85 (42.5)	107 (53.5)
	Number of runs	11 (5.5)	98 (49.0)	91 (45.5)

report. While the experiment proposed to the prospective teachers was a *generation task*, the production of the report includes a *recognition task*, since the subjects were asked to recognise the features of randomness in the data. According to Wagenaar (1972) and Falk and Konold (1997), these recognition tasks are more appropriate for assessing subjective perception of randomness since a person could have a good perception of randomness in spite of being unable to reproduce it.

Most participants represented the data using different graphs that varied in complexity and that have been analysed in a previous paper (Batanero et al. 2010). Many of them also computed averages (mean, median or modes) and variation parameters (range or standard deviation). Although the graphs and summaries produced were generally correct and many of them used Excel to produce a variety of statistical analyses, few of them got a completely correct conclusion about the group intuitions (Table 2).

Only a small number of prospective teachers explicitly were able to get a correct conclusion for the class' perception of expected values and variation in the different variables. Those who succeeded completed an informal inference process and were able to relate the empirical data to the problem posed in the project (the intuitions in the group), completing a modelling cycle (posing a problem; collecting data; using mathematical models, building with the model and interpreting the results as regards the problem posed). They concluded that the group had good intuitions as regards the average number of heads and, at the same time, that the perception of variability in this random variable was poor. An example is given below (similar responses were obtained for the other variables):

> *As regards the number of heads, the intuitions in the classroom were very close to what happens in reality; but not complete. The means in the real and simulated sequences are very close; the medians and modes are identical; however, the standard deviations suggest the spread in both distributions is different* (*CG*).

Other participants reached a partial conclusion, being able only to conclude about the central tendency or about the spread in the data. For example, the following teacher was able to perceive the similarity of means, but did not realise that the variation in the two sequences was quite different:

> *Observing the table, I think that my colleagues have good intuition since the most frequent values for the number of heads in the simulated sequence coincide with those in the real*

sequence; 10 and 11 are the most frequent values in both cases. The means are close to 10 in both sequences; therefore, the intuitions are good (*TG*).

In the next example, the student concluded about the difference in spread, but was unable to relate this result to the students' intuitions. He argued that the different students had similar intuitions but did not relate these intuitions to the results from the generation task experiment and did not analyse the central tendency:

Intuitions are very similar for the different students; there is not much irregularity. But when we compare to the real sequences, we realise the graph is more irregular. In the real sequence, the maximum number of heads is 16 and the minimum 7; however, in the simulated graph the maximum is 13 and the minimum 8; the range is smaller than that of real flipping of the coin (*MM*).

The remaining students either were unable to conclude or reached an incorrect conclusion. Part of them could not connect the results of their statistical analyses to the students' intuitions; that is, they did not see the implications of the results provided by the mathematical model to the solution of the problem posed (assessing the students' intuitions). An example is given below:

When I compare the data, I realise that many students agreed in their results. In spite of this, I still think there is mere chance since in the simulated sequences we invented the results (*EL*).

Other students connected the mathematical work to the problem situation, but they failed in their conclusions because they made an incorrect interpretation of the question posed by the lecturer. They assumed a good intuition would mean getting exactly the same results in the simulated and real sequences. In the next example, the prospective teacher shows a correct conception of randomness (randomness means lack of prediction in the short term) mixed with an incorrect conception: Instead of comparing the two distributions, he compared the students one by one and tried to assess the number of coincidences between the number of heads in the real and simulated sequences for each student:

Studying the graphs, the prediction of the group was not too bad. A game of chance is unpredictable; but when we count the number of students who guessed the result, the number of correct guesses is higher than the number of failures (*students who were very far from the real result*) (*LG*).

Although the binomial distribution (number of heads) was more intuitive for the prospective teachers, still the number of correct conclusions as regards the perception of the binomial distribution was very small. These results suggest that these prospective teachers did not only hold some misconceptions of randomness, they were also unaware of their misconceptions and were unable to recognise these misconceptions when confronted with the statistical data collected in the experiments.

5.5 Further Analysis of Pre-service Teachers' Conceptions

Another interesting point is that some of these teachers justified their wrong conclusions as regards some of the variables in the project by making explicit their own views of randomness that reproduced some of the conceptions described by Batanero and Serrano (1999) in secondary school students. Below we present examples of these conceptions, many of which are partly correct, but are incomplete and parallel some views of randomness that were described in Sect. 2.

Randomness and Causality The principle of cause and effect is deeply rooted in human experience and there is a tendency to relate randomness to causality. A possible way to relate randomness and cause is to think that what appears as randomness to a person's limited mind could well be explained by an extremely complex causal system that is unknown to the person, who is incapable of perceiving the causes for the phenomena. Another view is considering that causality is an illusion and that all events are actually random. Some participants assumed "Chance" as the cause of random phenomena, as was apparent in some participants' responses:

> *We define random experiments as a consequence of chance* (*SG*).
> *Some people think that the number of possibilities of getting a tail is 50 %; but, although in this experiment results were very close to what we expected, this result was due to chance or good luck because we cannot deduce this result, since it depends on chance* (*NG*).

Randomness as Unpredictability A common feature in different conceptions of randomness is unpredictability: the fact that we cannot predict a future event based on a past outcome (Bennett 1998). Some participants expressed this idea in their responses, since they assumed they could not reach a conclusion about the differences in distribution for the number of runs, number of heads or longest run because anything might happen in a random process. In these responses, the "outcome approach" (Konold 1989), that is, the interpretation of probability questions in a non-probabilistic way, may also operate:

> *I want to note that it is impossible to make a prediction of the results since in this type of experiment any result is unpredictable* (*AA*).
> *Results of random experiments cannot be predicted until they happen* (*SG*).
> *My final conclusion is that in experiences related to chance, there are more or less likely events, but it is impossible to predict the exact result* (*MN*).

Randomness as Equiprobability A few subjects connected randomness as equiprobability (in the classical approach to this concept) and stated that any result was possible since the experiment was random and, consequently, there was equal probability for each result. This view was parallel to the classical conception of probability where an event is random only in case where there is the same probability for this event and for any other possible event in the experiment.

> *The probability for heads and tails is the same, and therefore, in 20 throws there is the same probability to obtain 20 heads, 20 tails or any possible combination of heads and tails* (*EB*).
> *You know that there is a 50 % possibility to get heads and another 50 % possibility to get tails because there are only two different results. Consequently, each student got a different result* (*CG*).

Lack of Pattern or Lack of Order Some participants associated randomness to lack of model or lack of pattern, a view close to von Mises's (1952) modelling of randomness, where a sequence of outcomes is random whenever it is impossible to get an algorithm that serves to produce the sequence. In particular, some students rejected the idea that a random sequence could appear as ordered. In spite that, this view is partly correct, in fact, in the analysis of the project data a variety of models appeared, such as the binomial distribution, the distribution of runs, or the geometrical distribution. These models arise in any random sequence and, consequently, randomness should have also been interpreted as multiplicity of models.

You cannot find a pattern, as it is random (*BS*).
We do not think it is possible to get a sequence so well ordered as $[C, C, C, C, C, +, +, +, +, +]$, *since our intuition led us to alternate between heads and tails and to produce unordered sequences, such as, for example*, $[C, C, +, C, +, +, C, +, C, +, C, +]$ (*RE*).
It is not random, it is too ordered (*SG*).

Randomness and Control A few prospective teachers described randomness as something that cannot be controlled, a vision common until the Middle Ages according Bennett (1998):

Despite our inability to control randomness, we got equal number of heads and tails (*AG*).
These predictions are really accurate. Although the simulation and reality are very close, we should take into account that randomness can never be controlled 100 %, even if you have much knowledge of the situation (*EC*).

On the contrary, the illusion of control (Langer 1975) defined as the expectancy of a personal success probability inappropriately higher than the objective probability would warrant was also observed. Consequently of this belief, some participants believed they could predict or control the result of the experiment. For example, one participant classified all his classmates according to their capacity to predict the results:

Only 21.7 % of students guessed the number of heads in the experiment; 13 % were very close because they had an error of (± 1)*; the remaining students failed in their prediction* (*LG*).

6 Discussion and Implications for Training Teachers

As stated by Bar-Hillel and Wagenaar (1991), randomness is a concept which somehow eludes satisfactory definition; although theoretically randomness is a property of a generating (random) process, in practice we can only infer indirectly randomness from properties of the generator's outcomes. In addition, although expressions such as 'random experiments', 'random number', 'random variable', 'random event', 'randomness' frequently appear in daily language as well as in school textbooks, the meaning of randomness in not clarified in school textbooks, thus increasing the likelihood of students having difficulties with this point (Batanero et al. 1998). It is not then surprising that, given such complexity, the prospective teachers in our sample showed different misconceptions of randomness, in both the sequences they produced and in their reports when analysing their data.

However, understanding randomness is an essential step in learning probability and therefore it is essential that prospective teachers acquire a sound understanding of this concept along their initial education if we want them succeed in their future teaching of probability. As stated by Ball et al. (2001, p. 453), some of the activities in which teachers regularly engage, such as "figuring out what students know; choosing and managing representations of mathematical ideas; selecting and modifying textbooks; deciding among alternative courses of action" involve mathematical reasoning and thinking. Consequently, teachers' instructional decisions as regards the teaching of probability are dependent on the teacher's probabilistic knowledge.

Prospective teachers in our sample showed a mixture of correct and incorrect beliefs concerning randomness. On the one hand, their perception of averages in the binomial distribution was good since, as stated by Falk et al. (2009), one major characteristic of a sequence of coin tosses is the equiprobability of the two outcomes and equal proportions of the two symbol types are more likely to be obtained by chance than any other result.

However, at the same time, misconceptions related to variation, and independence, as well associating randomness with ignorance, or trying to control randomness, also appear in the sample. This is a cause for concern when paired with evidence that prospective mathematics teachers in our sample may have a weak understanding of randomness and present different biases that could be transmitted to their future students.

These prospective teachers in our research were asked to solve a problem and complete a modelling cycle. According to Chaput et al. (2011), modelling consists of describing an extra-mathematical problem in a common language and building up an experimental protocol in order to carry out an experiment (in this research, the problem consisted in checking the intuitions of randomness and a particular experiment was chosen to get data on the teachers' intuitions). This description leads to setting some hypotheses which are intended to simplify the situation (in the example, the length of the sequences to be produced was fixed and the equiprobability of heads and tails in the coin was assumed).

Next, the second step of the modelling process is translating the problem and the working hypotheses into a mathematical model in such a way that working with the model produces some possible solution to the initial problem. The teachers translated the question (what conceptions they had on randomness) and the working hypotheses to statistical terms (they compared three pairs of distributions: the number of heads, number of runs and length of the longest run in both sequences and for the whole classroom). Consequently, participants in our sample built and worked with different statistical models (each student chose and produced particular graphs, tables or statistical summaries to compare these pairs of distributions).

The third and final step consisted of interpreting the mathematical results and related these results to reality in such a way that they produced some answers to the original problem. Although the majority of participants in our research correctly completed steps 1 and 2 in the modelling cycle, few of them were capable of translating the statistical results they got to a response about what the intuitions of the classroom on randomness were like. That is, few of them could understand what the statistical results indicated about the intuitions in the group and, therefore, these prospective teachers failed to complete the last part of the modelling process.

Dantal (1997) suggests that in our classroom we concentrate on step 2 ("the real mathematics) in the modelling cycle since this is the easiest part to teach to our students. However, all the steps are equally relevant for modelling and in learning mathematics if we want our students to understand and appreciate the usefulness of mathematics. It is therefore very important that teachers' educators develop the prospective teachers' ability to model in probability and the capacity to learn from data if we want to succeed in implementing statistics education at the school level.

Considering the importance and difficulty of the topic, the question remains about how to make randomness understandable by prospective teachers. Possibly this understanding should be developed gradually, by experiments and simulations, starting with concrete materials and moving later to computer simulations and observation of randomness in demographic or social phenomena. As stated by Batanero et al. (1998), it is important that prospective teachers understand that, in randomness apparent disorder, a multitude of global regularities can be discovered. These regularities allow us to study random phenomena using the theory of probability.

Finally, we suggest the usefulness of working with activities similar to the one described in this report to help prospective teachers make the conceptions about randomness and probability explicit. In order to overcome possible misconceptions, after working with these activities, it is important to continue the formative cycle with a didactical analysis of the situation. In our experience, in the second session, the correct and incorrect solutions to the project "checking your intuitions on randomness" were debated and the different conceptions of randomness explicit in the teachers' responses were discussed.

Acknowledgements Research supported by the project: EDU2010-14947 (MCINN-FEDER) and Group FQM196 (Junta de Andalucía).

References

Ayer, A. J. (1974). El azar [Chance]. In M. Kline (Ed.), *Matemáticas en el mundo moderno* (pp. 171–172). Barcelona: Blume.

Azcárate, P., Cardeñoso, J. M., & Porlánd, R. (1998). Concepciones de futuros profesores de educación primaria sobre la noción de aleatoriedad [Prospective primary school teachers' conceptions on the notion of randomness]. *Enseñanza de las Ciencias*, *16*(1), 85–97.

Ball, D. L., Lubienski, S. T., & Mewborn, D. S. (2001). Research on teaching mathematics: the unsolved problem of teachers' mathematical knowledge. In V. Richardson (Ed.), *Handbook of research on teaching* (pp. 433–456). Washington: American Educational Research Association.

Bar-Hillel, M., & Wagenaar, W. A. (1991). The perception of randomness. *Advances in Applied Mathematics*, *12*, 428–454.

Batanero, C., Arteaga, P., & Ruiz, B. (2010). Statistical graphs produced by prospective teachers in comparing two distributions. In V. Durand-Guerrier, S. Soury-Lavergne, & F. Arzarello (Eds.), *Proceedings of the sixth European conference of research in mathematics education, CERME-6 conference*. Lyon: ERME. Online: www.inrp.fr/editions/editions-electroniques/cerme6/.

Batanero, C., & Díaz, C. (2012). Training teachers to teach probability: reflections and challenges. *Chilean Journal of Statistics*, *3*(1), 3–13.

Batanero, C., Godino, J. D., & Roa, R. (2004). Training teachers to teach probability. In *Journal of Statistics Education* (Vol. 12). Online: www.amstat.org/publications/jse/.

Batanero, C., Henry, M., & Parzysz, B. (2005). The nature of chance and probability. In G. A. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 15–37). New York: Springer.

Batanero, C., & Serrano, L. (1999). The meaning of randomness for secondary school students. *Journal for Research in Mathematics Education*, *30*(5), 558–567.

Batanero, C., Serrano, L., & Green, D. R. (1998). Randomness, its meanings and implications for teaching probability. *International Journal of Mathematical Education in Science and Technology*, *29*(1), 113–123.

Bennett, D. J. (1998). *Randomness*. Cambridge: Harvard University Press.

Borovcnik, M. (2011). Strengthening the role of probability within statistics curricula. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics—challenges for teaching and teacher education. A joint ICMI/IASE study* (pp. 71–83). New York: Springer.

Borovcnik, M. (2012). Multiple perspectives on the concept of conditional probability. *AIEM. Avances de Investigación en Educación Matemática*, *2*, 5–27.

Chadjipadelis, T., Meletiou, M., & Paparistodemou, E. (2010). Statistics teacher of the new era: another specialized mathematician or a totally different person? In C. Reading (Ed.), *Proceedings of the eight international conference on teaching statistics*. Lubjana: International Association for Statistical Education. Online: www.stat.auckland.ac.nz/~iase/publications/icots8.

Chaput, B., Girard, J. C., & Henry, M. (2011). Modeling and simulations in statistics education. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics—challenges for teaching and teacher education. A joint ICMI/IASE study* (pp. 85–95). New York: Springer.

Chernoff, E. (2009a). *Subjective probabilities derived from the perceived randomness of sequences of outcomes*. Unpublished Ph.D. Thesis, Simon Fraser University.

Chernoff, E. (2009b). Sample space partitions: an investigative lens. *The Journal of Mathematical Behavior*, *28*, 19–29.

Dantal, B. (1997). Les enjeux de la modélisation en probabilité. In M. Henry (Coord.), *Enseigner les probabilités au lycée* (pp. 57–59). Reims: Commission Inter-IREM (The betting of modelling in probability).

Engel, J., & Sedlmeier, P. (2005). On middle-school students' comprehension of randomness and chance variability in data. *ZDM. Zentralblatt für Didaktik der Mathematik*, *37*(3), 168–177.

Falk, R. (1981). The perception of randomness. In C. Laborde (Ed.), *Proceedings of the fifth international conference for the psychology of mathematics education* (pp. 222–229). Grenoble: University of Grenoble.

Falk, R., Falk, R., & Ayton, P. (2009). Subjective patterns of randomness and choice: some consequences of collective responses. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(1), 203–224.

Falk, R., & Konold, C. (1997). Making sense of randomness: implicit encoding as a basis for judgment. *Psychological Review*, *104*, 301–318.

Fine, T. L. (1973). *Theories of probability. An examination of foundations*. London: Academic Press.

Fischbein, E., & Gazit, A. (1984). Does the teaching of probability improve probabilistic intuitions? *Educational Studies in Mathematics*, *15*(1), 1–24.

Fischbein, E., Nello, M. S., & Marino, M. S. (1991). Factors affecting probabilistic judgements in children and adolescents. *Educational Studies in Mathematics*, *22*, 523–549.

Franklin, C., Kader, G., Mewborn, D. S., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2005). *A curriculum framework for K-12 statistics education* (GAISE report). American Statistical Association. Online www.amstat.org/education/gaise/.

Franklin, C., & Mewborn, D. (2006). The statistical education of pre-K-12 teachers: a shared responsibility. In G. Burrill (Ed.), *NCTM 2006 yearbook: thinking and reasoning with data and chance* (pp. 335–344). Reston: NCTM.

Gal, I. (2005). Towards "probability literacy" for all citizens: building blocks and instructional dilemas. In G. Jones (Ed.), *Exploring probability in schools: challenges for teaching and learning* (pp. 39–63). New York: Springer.

Godino, J. D., Batanero, C., Roa, R., & Wilhelmi, M. R. (2008). Assessing and developing pedagogical content and statistical knowledge of primary school teachers through project work. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE study: teaching statistics in school mathematics. Challenges for teaching and teacher education. Proceedings of the ICMI study 18 and 2008 IASE round table conference*. Monterrey: ICMI and IASE. Online: www.stat.auckland.ac.nz/~iase/publications.

Green, D. R. (1983). A survey of probabilistic concepts in 3000 students aged 11–16 years. In D. R. Grey et al. (Eds.), *Proceedings of the first international conference on teaching statistics (v.2)* (pp. 766–783). Sheffield: University of Sheffield.

Green, D. R. (1989). Schools students' understanding of randomness. In R. Morris (Ed.), *Studies in mathematics education: Vol. 7. The teaching of statistics* (pp. 27–39). Paris: Unesco.

Green, D. R. (1991). *A longitudinal study of pupil's probability concepts* (Technical Report ME90/91). Loughborough University of Technology.

Hawkins, A., Jolliffe, F., & Glickman, L. (1991). *Teaching statistical concepts*. London: Longman.

Harten, G., & Steinbring, H. (1983). Randomness and stochastic independence. On the relationship between intuitive and mathematical definition. In R. W. Scholz (Ed.), *Decision making under uncertainty* (pp. 363–373). Amsterdam: Reidel.

Jacobbe, T. (2010). Preparing elementary school teachers to teach statistics: an international dilemma. In C. Reading (Ed.), *Proceedings of the eight international conference on teaching statistics*. Ljubljana: International Association for Statistical Education. Online: www.stat.auckland.ac.nz/~iase/publications/icots8.

Jones, G. (2005). Introduction. In G. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 1–12). New York: Springer.

Johnston-Wilder, P., & Pratt, D. (2007). The relationship between local and global perspectives on randomness. In D. Pitta & G. Philippou (Eds.), *Proceedings of the fifth European conference on research in mathematics education*. Lamarca: ERME. Online: www.erme.unito.it/CERME5b/.

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

Konold, C., Lohmeier, J., Pollatsek, A., Well, A. D., Falk, R., & Lipson, A. (1991). Novice Views on Randomness. In R. G. Underhill (Ed.), *Proceedings of the thirteenth annual meeting of the North American chapter of the international group for the psychology of mathematics education* (pp. 167–173). Blacksburg: Virginia Polytechnic Institute and State University.

Kyburg, H. E. (1974). *The logical foundations of statistical inference*. Boston: Reidel.

Lahanier-Reuter, D. (1999). *Conceptions du hasard et enseignement des probabilités statistiques [Conceptions of chance and teaching of statistics and probability]*. París: Presses Universitaires de France.

Liu, Y., & Thompson, P. (2002). Randomness: rethinking the foundation of probability. In D. Mewborn, P. Sztajn, E. White, H. Wiegel, R. Bryant, & 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 (Vol. *3*, pp. 1331–1334). Columbus. Eric Clearinghouse.

Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology*, *32*, 311–328.

Laplace, P. S. (1995). *Théorie analytique des probabilités (Analytical theory of probabilities)*. Paris: Jacques Gabay (Original work published 1814).

Lecoutre, M. P., Rovira, K., Lecoutre, B., & Poitevineau, J. (2006). People's intuitions about randomness and probability: an empirical study. *Statistics Education Research Journal*, *5*, 20–35. Online: www.stat.auckland.ac.nz/~iase/serj/.

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

MEC (2006). *Real Decreto 1513/2006, de 7 de diciembre, por el que se establecen las enseñanzas mínimas correspondientes a la Educación primaria*. Madrid: Author.

Ministério da Educação (1997). *Parâmetros curriculares nacionais: Matemática [National curricular parameters Mathematics]*. Brasilia: Author.

NCTM (2000). *Principles and standards for school mathematics*. Reston: NCTM. Online: standards.nctm.org/.

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

Piaget, J., & Inhelder, B. (1951). *La genése de l'idée de hasard chez l'enfant [The origin of the idea of chance in children]*. Paris: Presses Universitaires de France.

Stohl, H. (2005). Probability in teacher education and development. In G. Jones (Ed.), *Exploring probability in schools: challenges for teaching and learning* (pp. 345–366). New York: Springer.

Poincaré, H. (1936). Chance. *Journal of the American Statistical Association*, *31*, 10–30.

Pratt, D. (2000). Making sense of the total of two dice. *Journal for Research in Mathematics Education*, *31*(5), 602–626.

Toohey, P. G. (1995). *Adolescent perceptions of the concept of randomness*. Unpublished Master's Thesis, University of Adelaide.

Tversky, A., & Kahneman, D. (1982). Judgement of and by representativeness. In D. Kahneman, P. Slovic y, & A. Tversky (Eds.), *Judgement under uncertainty: heuristics and biases* (pp. 84–98). Cambridge: Cambridge University Press.

von Mises, R. (1952). *Probability, statistics and truth*. London: William Hodge. (Original work published in 1928).

Wagenaar, W. A. (1972). Generation of random sequences by human subjects: a critical survey of literature. *Psychological Bulletin*, *77*, 65–72.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(3), 223–265.

Zabell, S. L. (1992). The quest for randomness and statistical applications. In F. Gordon & S. Gordon (Eds.), *Statistics for the XXI Century* (pp. 139–150). Washington: Math. Assoc. of America.

Challenges of Developing Coherent Probabilistic Reasoning: Rethinking Randomness and Probability from a Stochastic Perspective

Luis Saldanha and Yan Liu

Abstract The concept of probability plays a vital role in mathematics and scientific research, as well as in our everyday lives. It has also become one of the fastest growing segments of high school and college curricula, yet learning probability within school contexts has proved more difficult than many in education realize.

This chapter is in two broad parts. The first part synthesizes a discussion of randomness and probability that is situated at the nexus of bodies of literature concerned with the ontology of stochastic events and epistemology of probabilistic ideas held by people. Our synthesis foregrounds philosophical, mathematical, and psychological debates about the meaning of randomness and probability that highlight their deeply problematic nature, and therefore raises the equally problematic question of how instruction might support students' understanding of them. We propose an approach to the design of probability instruction that focuses on the development of coherent *meanings* of randomness and probability—that is, schemes composed of imagery and conceptual operations that stand to support students' coherent thinking and reasoning about situations that we see as entailing randomness and probability. The second part of the chapter reports on aspects of a sequence of classroom teaching experiments in high school that employed such an instructional approach. We draw on evidence from these experiments to highlight challenges in learning and teaching stochastic conceptions of probability. Our students' challenges centered on re-construing given situations as idealized random experiments involving the conceptualization of an unambiguous and essentially repeatable trial,

Research reported in this chapter was supported by National Science Foundation Grant No. REC-9811879. Any conclusions and recommendations stated here are those of the authors and do not necessarily reflect official positions of NSF. We wish to acknowledge the contribution of Patrick W. Thompson—dissertation advisor to both authors, and principal investigator of the research project on which this chapter is based.

L. Saldanha (✉)
Arizona State University, Tempe, AZ, USA
e-mail: Luis.Saldanha@asu.edu

Y. Liu
SERP Institute, Washington, DC, USA

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_20,

as a basis for conceiving of the probability of an event as its anticipated long-run relative frequency.

1 Introduction

The concept of probability plays a vital role in mathematics, in scientific research, and in philosophy. It is also assuming an increasingly greater role in everyday lives throughout our society. Major political, social, economic, and scientific decisions are made using information based on probability models. In turn, probability has become one of the fastest growing segments of the high school and college curricula. Learning probability within school contexts, however, has proved more difficult than many in education realize. The profound challenges for students are associated not only with acquiring new skills but also with overcoming misconceptions of probability (Kahneman and Tversky 1972; Konold 1989), making sense of probability as a mathematical model (Shaughnessy 1977, 1992; Biehler 1994), and dealing with pervasive traditional teaching that often works against efforts to make sense of probabilistic concepts and ideas (Fischbein 1975). All these challenges add to the complexities with which statistics education researchers and teachers must grapple when designing instruction and curricula that support students' learning of probability.

This chapter unfolds in two broad parts. We begin by synthesizing a discussion of randomness and probability that is situated at the nexus of bodies of literature concerned with the ontology of stochastic events and epistemology of probabilistic ideas held by people. Our synthesis brings to the foreground philosophical and mathematical debates about the meaning of randomness and probability that highlight their deeply problematic nature, and therefore raises the equally problematic question of how instruction might support students' understanding of them. Ideas of randomness and probability can be seen to range along an ontological and epistemological spectrum. On one end of the spectrum is a deterministic stance in which randomness and probability are regarded *not* as inherent features of objective nature, but rather as residing wholly in the mind as an expression of our ignorance of the causes of actions and therefore of the true deterministic course of events. On the other end of the spectrum, randomness and probability are seen as inherent features of nature in which "genuinely statistical sequences" are thought to exist (von Mises 1957). In our view, both ends of the spectrum presume that the question is ontological—that it is about what randomness and probability *are*—and approach it by speaking of them as properties of something that exists, be it a sequence, a process, or a distribution. We propose an approach to the design of probability instruction that focuses instead on the development of coherent *meanings* of randomness and probability, that is, a focus on the development of schemes composed of imagery and conceptual operations that stand to support students' coherent thinking and reasoning about situations that we see as entailing randomness and probability.

In the second part of the chapter, we report on aspects of a sequence of classroom teaching experiments that employed such an instructional approach. We draw on evidence from the experiments to highlight challenges in learning and teaching stochastic conceptions of probability. Our experiments engaged a group of high school students in construing given situations in terms of stochastic events (Saldanha and Thompson 2007; Saldanha 2011). This construal involves a salient image of sampling as a process that one can imagine repeating under essentially identical conditions, and for which the probability of an outcome of interest produced by that process is conceived as the outcome's anticipated long-run relative frequency. A central scheme of images and operations that we intended for students to develop within this meaning is of random sampling as a loosely-coupled process with imprecisely-determined inputs, where the process generates collections of sampling outcomes having predictable distributions in the long run but unpredictable ones in the short run. Our students' challenges centered on construing given situations as idealized random experiments involving the conceptualization of an unambiguous and essentially repeatable trial, as a basis for conceiving of the probability of an event as its anticipated long-run relative frequency.

2 A Historical Perspective on Theories of Probability

Psychological and educational studies have revealed students' difficulties in understanding the concept of probability. Some concur that evidence of these difficulties can be traced back to the historical development of probability theory (Hacking 1975). Probability, as a mathematical theory, did not emerge before the mid-seventeenth century. Historically, the concept was often used to refer to two kinds of knowledge: *frequency-type probability* "concerning itself with stochastic laws of chance processes," and *belief-type probability* "dedicated to assessing reasonable degrees of belief in propositions quite devoid of statistical background" (Hacking 1975, p. 12; Hacking 2001, pp. 132–133). Circa 1654, there was an explosion of conceptions in the mathematical community that were compatible with this dual concept of probability, for example, frequentist probability, subjective probability, axiomatic probability, and, probability as propensity (cf. Gillies 2000; von Plato 1994). Yet, to this day, mathematicians and scientists continue to debate and negotiate meanings of probability both for its theoretical implications, and for its applications in scientific research. There are subjectivists, e.g., de Finetti (1970), who argue that frequentist or objective probability can be made sense of only through personal probability. There are frequentists, e.g., von Mises (1957), who contend that frequentist concepts are the only ones that are mathematically viable. According to Hacking (1975), although most people who use probability in practice are largely unconcerned with such distinctions, extremists of both schools "argue vigorously that the distinction is a sham, for there is only one kind of probability" (ibid., p. 15).

The co-existence of these conflicting theories of probability does, however, pose a great challenge to the instruction of probability. The question of "What to educate?" is contentious (Nilsson 2003), and disagreement on what constitutes probability underlies differences in educational researchers' use of terminology in this area. As Hawkins and Kapadia (1984) observed, the research findings in this area "are prone to confusion over what sort of 'probability' concepts are being researched, and as a result the findings are far from being immediately applicable to the classroom" (ibid., p. 374). Shaughnessy (1992) concurred that the historical development and the philosophical controversies surrounding the theories of probability "have saddled us with considerable baggage, which can provide obstacles not only to our research in learning probability and statistics, but also to our ability to communicate results to other researchers" (ibid., p. 467). Against this background, we will first provide a brief overview of the theories of probability.

There are many different views about the nature of probability and its associated concepts. Fine (1973), von Plato (1994), Gillies (2000), Hendricks et al. (2001) provided overviews of the debates that had been ongoing since the early seventeenth century, and Todhunter (1949), David (1962), and Hacking (1975) provided overviews of the development of probability prior to that. In what follows, we discuss a representative set of interpretations of probability that have profoundly influenced the research and design of probability curricula and instruction thus far. The sequence of our discussion roughly follows the chronological order of the works reviewed, and aims to convey a sense of the historical development of probability theory.

2.1 Laplace's Classical Probability

The essential characteristic of classical, or Laplacian, probability is "the conversion of either complete ignorance or partial symmetric knowledge concerning which of a set of alternatives is true, into a uniform probability distribution over the alternatives." (Fine 1973, p. 167). The core of this approach is the "principle of indifference", in which alternatives are considered to be equally probable in the absence of known reasons to the contrary. For example, all outcomes are equally probable in tossing die or flipping a coin. The probability of the occurrence of any outcome is one out of the number of all possible outcomes. This approach was the most prevalent method in the early development of probability theory, as the origins of probability theory were games of chance involving the notion of equal possibilities of the outcomes supposed to be known a priori (Todhunter 1949; David 1962).

However, classical probability builds on the problematic assumption of an equal likelihood of alternative outcomes. Yet "equal likelihood" is exactly synonymous with "equal probability." This gave rise to von Mises' (1957) criticism that "unless we consider the classical definition of probability to be a vicious circle, this definition means the reduction of all distribution to the simpler case of uniform distribution" (ibid., p. 68).

2.2 Von Mises' Limiting Relative Frequency Probability

von Mises' (1957) relative frequency definition of probability was based on two central constructs: collective and randomness. Probability is only applied to infinite sequences of uniform events or processes that differ by certain observable attributes, which he labeled "the collective." Two hypotheses about collectives were essential in von Mises' theory of probability. The first was the existence of the limiting value of the relative frequency of the observed attribute. In other words, a collective must be a long sequence of observations for which the relative frequency of the observed attribute would tend to a fixed limit if the observations were indefinitely continued. The second hypothesis was a condition of randomness, which he characterized as "the impossibility of devising a method of *selecting the elements* so as to produce a fundamental change in the relative frequencies" (ibid., p. 24).

The strength of von Mises' limiting relative frequency theory of probability is that it offered both a physical interpretation of, and a way of measuring, probability. It offered an operational definition of probability based on the observable concept of frequency.[1] Objections to von Mises' theory attack its disconnection between theory and observation by the use of limits in infinite sequences (Gillies 2000). Two sequences can agree at the first n places for any finite n however large, and yet converge to quite different limits. Suppose a coin is tossed 1,000 times and that the observed frequency of heads is approximately $1/2$. This is compatible with the limit being quite different from $1/2$; in other words, the observation does not exclude the possibility that the probability is, say, 0.5007. Fine (1973) harmonized this criticism by adding that knowing the value of the limit without knowing how it was approached would not assist in making inferences, and therefore a limit interpretation is of value neither for the measurement of probability nor for the application of probability.

2.3 Kolmogorov's Measure-Theoretic Probability

Kolmogorov (1950) constructed a formal mathematical theory of probability on the basis of measure theory. In this theory, a probability space is defined as a triplet, (Ω, F, P), comprising a sample space, Ω; a σ-field F of selected subsets of Ω; and a probability measure or assignment, P. The sample space Ω contains elements called "elementary events," and the σ-field of subsets of Ω, F, satisfies properties that include closure under the operations of complementation and countable unions. The probability measure P is a nonnegative set function that maps elements of F to the interval $[0, 1]$. P satisfies the now familiar axioms which include that the

[1] Because of his background as a physicist, von Mises was more concerned with the link between probability theory and natural phenomena. von Mises regarded probability as "a *scientific theory* of the same kind as any other branch of the exact *natural science*" (ibid., p. 7).

measure of the entire sample space is equal to 1 ($P(\Omega) = 1$), and that the measure of any finite union of pairwise disjoint elements in F is equal to the finite sum of the measure of the individual elements ($P(\bigcup_{i=1}^{n} F_i) = \sum_{i=1}^{n} P(F_i)$, for pairwise disjoint $F_i \in F$).

Kolmogorov's approach to probability has been regarded as a benchmark in the development of formal probability theory. It is considered to be almost universally applicable in situations having to do with chance and uncertainty. This approach to probability made a distinction between probability as a deductive structure and probability as a descriptive science.[2] Kolmogorov's axiomatic approach had been difficult to accept by experimental probabilists: "The idea that a mathematical random variable is simply a function, with no romantic connotation, seemed rather humiliating..." (Doob 1996, p. 593). Fine (1973) argued that the probability scale in Kolmogorov's approach was "occasionally empirically meaningless and always embodies an arbitrary choice or convention" (ibid., p. 59).

2.4 De Finetti's Subjective Probability

De Finetti (1970) defined subjective probability as "the *degree of belief* in the occurrence of an event attributed by a given person at a given instant and with a given set of information" (ibid., p. 3). De Finetti believed that this definition of probability made no philosophical assumptions about probability. However, it is understood by many that the theory of subjective probability makes the assumption that probability is a degree of belief or intensity of conviction that resides within a human mind, as opposed to an objective probability that exists irrespective of mind and logic (Good 1965, p. 6).

The defining property of subjective probability is the use of further experiences and evidence to change the initial opinions or assignment of probability. This is expressed formally as Bayes' theorem. Good (1965) argued that it was not a belief in Bayes' theorem that made one a Bayesian, as the theorem itself was just a trivial consequence of the product axiom of probability. Rather, it was a readiness to incorporate intuitive probability into statistical theory and practices that made one a subjectivist. However, the very "subjective" character of de Finetti's approach has been the most intensively discussed and criticized. In particular, his intention to view probability as subjective was considered as introducing arbitrariness into probability theory, which "invalidates the power of Bayesian theory" (Piccinato 1986, p. 15).

---

[2] A probability curriculum starting with the axiomatic approach to probability may prevent students from developing an understanding of probability that is rooted in their experience. It is like teaching the concept of circle as *a set of points* (x, y) *that satisfies the condition* $(x - a)^2 + (y - b)^2 = c^2$, $a, b, c, \in \mathbb{R}$, $c \geq 0$. If this definition precedes the introduction of the concept of distance and measurement of length, students may experience difficulties in visualizing the image of a circle as a set of points having the same distance from a fixed point in a two-dimensional surface.

2.5 Discussion

Gillies (2000) divided the interpretations of probability into two types—*epistemological* and *objective*. Epistemological interpretations, such as subjective theory, take probability to be concerned with knowledge or beliefs of human beings, whereas objective interpretations including the relative frequency theory take it to be a feature of the world outside the human mind. However, from our point of view, both types are essentially concerned with ontology, in the sense that the discussion is focused on "what probability *is*" and "*where* it is located," be it beliefs or knowledge in a human mind or an object existing independent from human perception. In our view, a truly epistemological perspective would focus on what people *mean* by "probability." Hacking (1975) suggested two ways of studying the historical development of probabilistic thinking, which, for convenience, we label *a discourse perspective* and *a historical perspective*. A *discourse perspective* takes probability as existing "in the discourse and not in the minds of speakers." In this perspective, one is concerned "not with who wrote, but with what was said." A *historical perspective*, on the other hand, considers "how an idea is communicated from one thinker to another, what new is added, what error is deleted" (ibid., p. 16). From a discourse perspective, each theory of probability must be viewed as taken by its proponent as the one true understanding of what probability is. The history of probability is filled with vigorous debates about the only viable and true meaning of probability. In contrast, a person taking a historical perspective considers the historical contexts from which each theory emerges, the philosophical assumptions, and the genetic aspects hidden in each theory. We argue that, as a consequence of the historical approach, theories of probability can be viewed as valid models that emerged out of particular contexts, satisfied particular conditions, served particular purposes, and thus were not necessarily commensurable. When one considers making the grounds of various mathematical approaches visible and accessible for learners, then not all approaches are assumed to be equally suitable—some approaches might provide better starting points than others. This point of view is compatible with Shaughnessy's (1977, 1992) modeling perspective, wherein the meaning of probability one employs in a particular situation should be determined by the tasks and the types of problems that are involved. From this perspective, designing a probability curriculum and instruction is no longer a task of deciding which side to take, classical probability or modern axiomatic probability, subjective or objective, and so on. Rather, instructional design becomes an enterprise in which one considers the strengths and weaknesses of each theory, and elicits useful constructs and ways of thinking in designing the instructional objectives.

Although it is beneficial for educators to know the different theories of probability in order to develop a perspective that allows them to reconcile the differences among the theories and to avoid the potential pitfalls in instruction and research, designing instructional objectives takes more than that. A modeling perspective orients one to think about the instructional design of probability, but it does not directly address the question of *what* we want students to learn—as in what *understandings* and *meanings* we intend for them to develop. What does it mean to reason probabilistically? How might we describe probabilistic reasoning in its fine details so as to better guide students' learning? Addressing these questions in a serious manner necessitates consideration of the conceptual and psychological aspects of learning probability—the underlying beliefs, implicit presumptions, and informal knowledge people employ in reasoning about uncertainty, and our conjectures about developmental and conceptual trajectories people might follow in developing probabilistic reasoning. The next section of our chapter endeavors to consider such aspects.

3 Insight into Conceptions of Probability

3.1 Developmental Trajectories

In the previous section, we discussed the literature that addressed the question of what probability is. Similar to any general discussion on the nature of a scientific notion, this endeavor has too often fallen into philosophical quandaries. The research program of Genetic Epistemology, led by Jean Piaget, aimed to sidestep this philosophical entanglement by empirically pursuing the question of how one develops operations of thought that enable probabilistic reasoning.

Piaget characterized chance as an essential aspect of irreversible phenomena, as opposed to mechanical causality or determinism characterized by its intrinsic reversibility. In their seminal book *The origin of the idea of chance in children*, Piaget and Inhelder (1951/1975) described children's construction of chance and probability in relation to the development of their conceptual operations. According to Piaget and Inhelder, children develop the concepts of probability in three successive stages, which correspond largely with those of the evolution of logical-arithmetical and combinatorial operations.

In the first stage, children do not distinguish possible from necessary events. In the second stage, children begin to differentiate between the necessary and the possible through the construction of the concrete logical-arithmetical operations.[3] At this level, the notion of chance acquires a meaning as a "noncomposable and irreversible reality" (ibid., p. 223). Piaget and Inhelder further hypothesized that beyond the recognition of the clear opposition between the operative and the chance, the concept of probability presupposes the idea of a sample space, i.e., all the possible outcomes, so that "each isolated case acquires a probability expressed as a fraction of the whole" (ibid., p. 229). To get to this stage, children must be able to construct combinatorial operations and understand proportionalities. Piaget and Inhelder found, however, that after distinguishing the possible from the necessary,

[3]The failure of differentiating between possible and necessary as a developmental constraint was confirmed by later studies of Kahneman and Tversky (1982) and Konold (1989). They found that even when people overcame the developmental constraint, they might still fail, or rather refuse, to distinguish uncertain events from necessary ones due to a deterministic world view.

children of the second stage failed to produce an exhaustive analysis of the possible. They argued that this was because an analysis of sample space assumed operating on simple possibilities as *hypotheses*, whereas children at this stage were only able to deal with actual possibilities.

Finally, at the third stage, children translate unpredictable and incomprehensible chance into a system of operations that are still incomplete and effected unsystematically. As such, chance becomes comparable to those very operations conducted completely and systematically. Once children learn the operations of permutations, they can deduce all the possibilities of a chance situation and imagine one outcome occupying a relative portion of all possible outcomes. They can then imagine events (sets of possible outcomes) as having a relative size in comparison to all possibilities. These operations lead to the determination of all the possible cases, even though each of them remains indeterminate for its particular realization. Piaget's insistence that we could not characterize children's thinking about chance as probabilistic until they develop concrete mental operations was because he was trying to capture the feeling of *relative significance of an event in relation to a sense of the weight of all possibilities*, and that children cannot develop a sense of weight of all possibilities until they can imagine a systematic way of generating them (hence the need for combinatorial reasoning).

The most well known studies of students' conceptions of probability in the post-Piagetian era were conducted by Green (1979, 1983, 1987, 1989), and by Fischbein and his colleagues (Fischbein 1975; Fischbein and Gazit 1984; Fischbein et al. 1991; Fischbein and Schnarch 1997). Green surveyed over 3,000 students of 11 to 16 years of age in England to determine their levels of development and knowledge of probability concepts. He found that the students had a weak understanding of the concept of ratio and that many did not understand the common language of probability, such as "at least" or "certain," or "impossible." Following the Piagetian framework, Fischbein and colleagues investigated the "probabilistic intuition" and its evolution in relation to children's cognitive development. According to Fischbein, an intuition is a cognitive belief that the person having it feels is self-evident, self-consistent and hardly questionable. For instance, the statement *through a point outside a line one and only one parallel may be drawn to that line* is an intuition. An intuition may or may not contradict scientifically accepted knowledge. A probability intuition thus referred to a global, synthetic and non-explicitly justified evaluation or prediction (Fischbein and Gazit 1984). Fischbein identified the primary intuition of chance as "the opposite of 'determined' " (ibid., p. 70). Contrary to Piaget and Inhelder's claim that children did not differentiate chance and necessity before the concrete operational stage (around the age of 7–8), Fischbein argued that the intuition of chance might be identified in preschool children, i.e., at the pre-operational level. Fischbein's later research attempted to reveal the complexity and the varieties of misconceptions and biases one had to overcome in order to develop a coherent intuitive background for probabilistic reasoning (Fischbein et al. 1991). Such misconceptions, he found, were likely to be caused by linguistic difficulties (e.g., lack of a clear definition of the terms "possible," "impossible," and "certain"), lack of logical abilities, and the difficulty of extracting mathematical structure from the practical embodiment.

3.2 Judgment Heuristics

The research of Kahneman and Tversky (1972, 1973, 1982) documented the persistent "misconceptions" that people demonstrated when making judgments under conditions involving uncertainty. Among these misconceptions were systematic mental heuristics that do not conform to the mathematically normative ways of reasoning. For example, according to the "representativeness heuristic" people estimated the likelihood of an event based on how well an outcome represented some aspect of its parent population, or how well it reflected the process by which it was generated (Kahneman and Tversky 1972, p. 430). One problem often cited in the literature to illustrate the representativeness heuristic has been referred to as the "Linda Problem":

> Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Which of two statements about Linda was more probable: (i) Linda is a bank teller, or (ii) Linda is a bank teller who is active in the feminist movement (p. 297).

Eighty six percent of subjects that were given the problem chose statement (ii)—a choice that violated the conjunction rule of probability. Kahneman and Tversky attributed this violation to the subjects' having based their choice on its resemblance to the sketch. They thus concluded that the subjects did not have valid intuitions corresponding to the formal rules of probability when making judgments. Another example illustrates the "base-rate misconception":

> A cab was involved in a hit and run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data: 85 % of the cabs in the city are Green and 15 % are Blue. A witness identified the cab as Blue. The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colors 80 % of the time and failed 20 % of the time. What is the probability that the cab involved in the accident was Blue rather than Green? (Kahneman and Tversky 1982, pp. 156–157)

The correct answer is 41 %. The typical answer from a large number of subjects was 80 %, which coincided with the witness' accuracy rate. Kahneman and Tversky hypothesized that people tended to ignore the base rate information because they saw it as incidental.

Such judgment heuristics have been explained in two ways. One explanation concerns mathematics, and it says that some of the principles of mathematical probability are non-intuitive or counter-intuitive, and therefore account for subjects' difficulties. The second explanation concerns psychology; it claims that human minds are simply not built to work by the rules of probability (Gould 1992, p. 469; Piatelli-Palmarini 1994).

Later research by Konold and his colleagues (Konold 1989; Konold et al. 1993) and by Gigerenzer (1994, 1996, 1998), Hertwig and Gigerenzer (1999) suggested that Kahneman and Tversky might have over-interpreted their data. While Kahneman and Tversky focused on subjects' assessment or assignment of probability, both

Konold and Gigerenzer shifted the focus towards students' *interpretation* of probability. Konold (1989) wrote:

> Hidden in the heuristic account is the assumption that regardless of whether one uses a heuristic or the formal methods of probability theory, the individual perceives the goal as arriving at the probability of the event in question. While the derived probability value may be non-normative, the meaning of that probability is assumed to lie somewhere in the range of acceptable interpretation (ibid., p. 146).

In other words, Kahneman and Tversky seemed to assume that their subjects had a mathematical meaning for "probability", while it could have been that many of the subjects had a deterministic understanding of events, and that numerical probability simply reflected their degree of belief in the outcome.[4]

In their experiments employing the Linda problem, Hertwig and Gigerenzer (1999) found that many subjects had *nonmathematical* interpretations of "probability." Subjects perceived the problem as a task of looking for a plausible or accurate description of Linda. Such discrepancy in the interpretations of probability came from the mismatch of intentions: While the subjects assumed that the content of the Linda problem should be relevant to the answer, Kahneman and Tversky were in fact testing for sound probabilistic reasoning according to which the content is irrelevant, i.e., "all that counts are the terms *probable* and *and*" (Gigerenzer 1996, p. 593, italics in original). Hertwig and Gigerenzer (1999) contend that if subjects assumed a nonmathematical interpretation of probability, then their answers could not be taken as evidences of violation of probability theory "because mathematical probability is not being assessed" (ibid., p. 278).

3.3 The Outcome Approach

In his study of students' informal conceptions of probability, Konold (1989) found that students' preference for causal over stochastic models was linked to a tendency to predict outcomes of single trials. He termed this model of students' reasoning the *outcome approach*. Outcome oriented thinking was characterized by three salient features: (i) predicting outcomes of single trials, (ii) interpreting probability as predictions and thus evaluating probabilities as either right or wrong after a single occurrence, (iii) basing probability estimates on causal features rather than distributional information. Individuals employing an outcome approach did not interpret probability questions as having to do with a stochastic process. Instead of conceiving

[4]We note that there are two other potential sources of deviation from the standard solution for the taxi problem. If a subject conceptualizes the problem as $P(\textit{Taxi is Blue}|\textit{Witness says "blue"})$, then to give the standard solution, the subject must take the witness' accuracy rate as his or her propensity to say "Blue," and the subject might see no compelling reason to assume this because there is no information on the witness' propensity to say "I'm unsure". Second, if the subject takes it as a fact that the witness said "Blue", then $P(\textit{Witness says "blue"})$ is 1, and hence $P(\textit{Taxi is Blue}|\textit{Witness says "blue"})$ is equal to $P(\textit{Taxi is Blue})$.

a single trial or event as embedded within a sample of many such trials, they viewed each one as a separate, individual phenomenon. Consequently, they tended to interpret their decision-making task as one of correctly predicting for certain, and on the basis of relevant causal factors, what the next outcome would be, rather than one of estimating what is likely to occur over the long run on the basis of frequency data. Konold's finding suggested that deterministic reasoning was tied to students' understanding of the goal of probability as predicting the outcome. He further claimed that if the outcome approach was a valid description of some novices' orientation to uncertainty, then the application of a causal rather than a black-box model to uncertainty seemed to be the most profound difference between those novices and the probability expert and, therefore, perhaps the most important notion to address in instruction.

Konold (1995) also found that students could hold multiple, and sometimes contradictory beliefs about chance situations. In one experiment, students were given the following problems:

> Part 1: Which of the following sequences is most likely to result from flipping a fair coin 5 times? (a) H H H T T, (b) T H H T H, (c) T H T T T, (d) H T H T H, and (e) All four sequences are equally likely;
> Part 2: Which of the above sequences is least likely to result from flipping a fair coin 5 times?

Konold reported that while 70 % of the subjects correctly responded to Part 1 of the problem, that the sequences were equally likely, over half of the same subjects did not choose (e) for Part 2. These subjects indicated that one of the sequences was "least likely," which contradicted their response to the first part. After interviewing these subjects, Konold concluded that this inconsistency resulted from the subjects' applying different perspectives to the two parts of the problem. In Part 1, many subjects thought they were being asked, in accordance with the outcome approach, to predict which sequence would occur. They chose (e) not because they understood that the probability of each sequence was the same, but because they could not rule out any of them. In Part 2, many of these subjects applied the representativeness heuristics. For example, one might choose (c) as being the least likely based on a perception that it contains an excess of the letter T.

3.4 Deterministic Reasoning

In the context of investigating students' difficulties in understanding sampling, Schwartz et al. (Schwartz and Goldman 1996; Schwart et al. 1998) found that understanding sampling required the ability to manage the tensions between ideas of causality and randomness. For instance, understanding a public opinion poll as a random sample involved giving up analysis of the causal factors behind people's opinions. Schwartz et al. referred to people's tendency to focus on causal association in chance situations as the *covariance assumption*, which described specifically the phenomena that (i) people reason as though they assume everyday events should be explained causally, and (ii) people search for co-occurrences or associations between events and/or properties that can support this kind of explanation.

Biehler (1994) differentiated between what he called two *cultures of thinking*: exploratory data analysis and probabilistic thinking. He argued that while probabilistic thinking required ruling out deterministic reasoning, exploratory data analysis values seeking and interpreting connections among events. The inherent conflict between these two ways of thinking raised the question: Do we need a probabilistic revolution[5] after having taught data analysis? (ibid). While some researchers (Fischbein 1975; Moore 1990; Falk and Konold 1992; Metz 1998) concurred that in learning probability, students must undergo a similar revolution in their thinking, Biehler argued for an epistemological irreducibility of chance instead of an ontological indeterminism that the probabilistic revolution seemed to suggest (e.g., quantum mechanics as the epitome of an inherently non-deterministic view of natural phenomena). He said:

> … this ontological indeterminism, the concept of the irreducibility of chance is a much stronger attitude … the essence of the probabilistic revolution was the recognition that in several cases probability models are useful types of models that represent kinds of knowledge that would still be useful *even when further previously hidden variables were known and insights about causal mechanisms are possible*. (Biehler 1994, p. 4, italics in original)

This point of view resonates with Konold (1989), who suggested that a probability approach adopts a "black-box" model that ignores, if not denies, underlying causal mechanisms.

In summary, an epistemological stance or world view that one embraces as a general principle in guiding one's perception and actions is thought of as having to do with one's development of probabilistic reasoning. One who regards the world as being intrinsically deterministic may naturally reason deterministically when judging probabilities, whereas one who views the world as being non-deterministic will seek models and modeling in achieving maximum information about uncertain events. But thus far researchers seem to have disagreed on whether a change in a deterministic world view is necessary for learning probability. On one side of the debate, probabilistic reasoning is viewed as presupposing a "probabilistic revolution" in the mind. On the other side, probabilistic reasoning does not necessarily conflict with a deterministic world view; one may view the world as being connected through cause and effect, and yet intentionally avoid seeking causal factors and explanations. In the latter case, probability is considered as a model that is chosen for a certain situation with the purpose of approximating phenomena and gaining information, a choice founded upon a realization that the model is expected to be a more powerful predictor of outcomes over the long run than deterministic analyses. However, researchers do tend to agree that students having a strong deterministic world view are more likely to have a non-stochastic conception of events.

---

[5]The term "probabilistic revolution" (Krüger 1987) broadly suggests a shift in world view within the scientific community from 1800 through 1930, from a deterministic reality where everything in the world is connected by necessity in the form of cause–effect relationships, to one in which uncertainty and probability have become central and indispensable.

3.5 Proportional Reasoning

Early developmental studies (Piaget and Inhelder 1951/1975; Green 1979, 1983, 1987, 1989) demonstrated that weak proportional reasoning imposed limitations on children's ability to make probabilistic judgments. More recent studies (Fischbein and Gazit 1984; Garfield and Ahlgren 1988; Ritson 1998) confirmed that the ability to reason probabilistically was highly dependent on one's understanding of fractional quantities, ratios and proportions. This relationship between probability and proportionality might lead one to conclude that if one understood probability, one must have understood the concept of proportionality. Fischbein and Gazit (1984) found that this was not the case. They found that although probabilistic thinking and proportional reasoning shared the same root, which they called the *intuition of relative frequency*, the two were based on distinct mental schemata, and progress obtained in one direction did not imply an improvement in the other.

Ritson (1998) suggested that teachers should consider using probability as a context in which to teach the concepts of fraction and ratio. Evolutionary psychologists led by Gigerenzer (1998), on the contrary, proposed eliminating proportionality from probability instruction. He argued that relative frequencies and percentages are not natural ways of reasoning, and that instead the human mind is predisposed to attend to numerical information in uncertain situations in the form of natural frequencies. Gigerenzer (1994) and Hertwig and Gigerenzer (1999) showed that students ceased to employ judgment heuristics when they worked on problems formatted in natural frequencies. Sedlmeier (1999) demonstrated that natural frequencies proved effective in training people how to make probability judgments and Bayesian inferences. Despite these reported successes, we view the attempt to replace relative frequency with natural frequency as somewhat problematic. Gigerenzer (1998) suggested that teachers instruct people how to represent probability information in natural frequencies. Such ability ostensibly entails understanding of fractions, as most probability information that people encounter in their lives is expressed as fractions, ratios, or percentages. Moreover, in order to make a probability judgment from information presented in natural frequencies, one needs to understand the proportional relationship of the quantities that are involved.[6]

In summary, proportional reasoning and its related conceptual operations appear to support probabilistic judgment once students conceptualize a probabilistic event

[6]Consider the example given by Gigerenzer (ibid., p. 17): A scenario in probability format: The probability that a person has colon cancer is 0.3 %. If a person has colon cancer, the probability that the test is positive is 50 %; if not, the probability that the test is positive is 3 %. What is the probability that a person who tests positive actually has colon cancer? In natural frequency format, 30 out of every 10,000 people have colon cancer. Of these 30 people, 15 will test positive. Of the remaining 9,970 people, 300 will still test positive. To solve the problem in probability format, one uses Bayes' theorem: $(0.3\ \% \times 50\ \%)/(0.3\ \% \times 50\ \% + 99.7\ \% \times 3\ \%)$. In the natural frequency format, $15/(300 + 15)$ will suffice. However, to justify substituting numbers in place of the percentages requires understanding that the underlying proportional relationships will remain constant no matter the population's size.

as being a particular case in reference to a class of events. Although extensive research in mathematics education (Harel and Confrey 1994) indicates that the concepts of fraction, ratio, percentages, and other conceptions involving relative comparison of quantities are arguably no less complicated than that of probability, the role of proportional reasoning in the learning and teaching of probability has not been extensively researched.

3.6 Stochastic Conception of an Event: A Synthesis

In summary, developing probabilistic reasoning entails overcoming challenges that include developmental constraints, deterministic reasoning, outcome approach, judgment heuristics, holding multiple yet conflicting perspectives, and developing proportional reasoning. An overarching finding that we draw from looking across these challenges is a crucial distinction between two different modes of judgment that people adopt in attributing uncertainty. Kahneman and Tversky (1979) referred to these as a *distributional* mode and a *singular* mode. In a distributional mode, the particular case in question is seen as an instance of a class of similar cases, for which the relative frequencies of outcomes are known or can be estimated. In a singular mode, probabilities are assessed by the propensities of the particular case at hand. This distinction is central to our understanding of probabilistic reasoning. A student adopting a singular mode is likely to interpret the task of probability judgment as one of predicting the outcome of a single event (outcome approach), or to look for causal factors when searching for justifications for her judgment (deterministic reasoning), or to apply judgment heuristics. In contrast, a student adopting a distributional mode sees an event as one case of a class consisting of similar cases, and may consequently be likely to ignore the causal mechanisms and instead adopt a frequency approach in evaluating probability.

While acknowledging that many questions can be approached from either a singular or distributional mode, Kahneman and Tversky conjectured that people generally preferred the singular mode, in which they took an "inside view" of the causal system that most immediately produced the outcome, over an "outside view", which related the case at hand to a sampling scheme. The question arises as to how we might help students abandon a singular mode and adopt a distributional mode. How might a student conceive of one event as being a case from a given class?[7] This question is reminiscent of von Mises' idea of "collective" (von Mises 1957, p. 18). A collective is given as a priori, and is imposed on the probability model. This approach invited controversies about the justification of collectives and the existence

[7] We relate an anecdote in anticipation of Bayesians' objections to this statement: In a recent conversation with a prominent Bayesian probabilist, we asked "Suppose, at a given moment, and for whatever reason, you judge that the probability of an event is 0.4. At that moment, what does "0.4" mean to you?" He responded that it meant that he anticipates that this event will occur 40 % of the time.

of limit. To avoid such difficulty, we advocate a shift of focus from collectives to random processes from which collectives are generated. Liu and Thompson (2007) extended the distinction of singular and distributional mode to that of a *stochastic* versus *non-stochastic conception* of events. A stochastic conception of an event entails thinking of it as an instantiation of an underlying repeatable process, whereas a non-stochastic conception entails thinking of an event as unrepeatable or never to be repeated. To develop a stochastic conception, one has to go through a series of conceptions and operations:

- Thinking of a situation (an "event");
- Seeing the event as an expression of some process;
- Taking for granted that the process could be repeated under similar conditions;
- Taking for granted that the conditions and implementation of the process will differ among repetitions in small, yet perhaps important, ways;
- Anticipating that repeating the process would produce a collection of outcomes;
- And, reciprocally, seeing a collection as having been generated by a stochastic process.

This hypothetical trajectory brings to the fore a conception of random process, through which the concepts of events, outcomes, and collections of outcomes are interconnected as a coherent scheme of ideas.

4 Challenges in Developing Stochastic Conceptions

In this section, we draw on data from a sequence of classroom teaching experiments to illustrate some key challenges in learning and teaching the conceptions and operations entailed in developing a stochastic conception of event as described above.

4.1 Background

Our data is drawn from a larger research project that employed classroom teaching experiments as a method for investigating the possibilities and challenges of supporting high school students in developing coherent understandings of statistics and probability (Saldanha 2004; Saldanha and Thompson 2002, 2007).[8] The method entailed documenting and modeling students' thinking and conceptions related to probability as they emerged in relation to students' engagement with instruction designed to support such thinking. The data we present consists of excerpts from whole-class discussions and individual interviews with a group of eight statistically naïve high school students who participated in our experiments, which were conducted in an introductory statistics and probability class. We conducted a sequence

[8]The authors were members of the research team that conducted the project.

of two experiments that addressed probabilistic ideas per se, although we only introduced the term "probability" in the second experiment.

The first experiment focused on building the ideas of sampling distributions and sampling as a random process—one that can be repeated under essentially identical conditions and that produces unpredictable outcomes from trial-to-trial, but which gives rise to predictable patterns of outcomes over the long run. The experiment began by first engaging students with hands-on sampling activities wherein they constructed empirical sampling distributions of simple statistics (e.g., sample proportions). We then engaged students in designing (sampling) simulations of given situations that were not evidently stochastic, and we had them run the computer simulations that generated distributions of a sample statistic. Students studied patterns of dispersion within such distributions. They used those patterns as a basis for making relative-frequency-based judgments about the types of sampling outcomes one might expect over the long run, and about the relative unusualness of given outcomes. The data we present from the first experiment is drawn from the class discussions around the simulation design activities.[9]

The second experiment built directly on students' experiences in the first experiment by moving to support them in developing a long-run relative-frequency meaning for probability. The data we present from the second experiment is drawn from interviews conducted at the conclusion of the experiment's first phase; instruction in that phase focused on having students practice identifying whether given situations were "probabilistic" per se—that is, involving a repeatable stochastic experiment—and then re-construing and describing such situations in probabilistic terms when possible.

4.2 Construing Situations as Stochastic Experiments

The excerpts we discuss here are drawn from the classroom discussions around the task displayed in Fig. 1, wherein students were to design a simulation of the "Movie Theatre" situation[10] with the constraints of the Prob Sim (Konold and Miller 1996) micro-world interface in mind. Prob Sim's interface employed the metaphor of a "mixer" for a population and a "bin" for a sample. It allowed the user to easily specify a population's composition and size, the size and selection method of a sample (with or without replacement), and the number of trials of the simulated sampling experiment to be conducted. Students had already used the software in the preceding phase of the teaching experiment.

The discussions around the movie theatre situation brought to light difficulties that students experienced in reconceiving a given situation as an idealized stochastic experiment. The task entailed construing and re-describing the given situation in

[9] Part of this data was previously reported in Saldanha (2010, 2011).

[10] Adapted from Konold (2002).

Investigating "Unusualness"

Ephram works at a theatre, taking tickets for one movie per night at a theatre that holds 250 people. The town has 30, 000 people. He estimates that he knows 300 of them by name.

Ephram noticed that he often saw at least two people he knew. Is it in fact unusual that at least two people Ephram knows attend the movie he shows, or could people be coming because he is there?
(The theatre holds 250 people)

Assumptions for your investigation:

Method of investigation:

"Gut level" answer:

Result:

Conclusion:

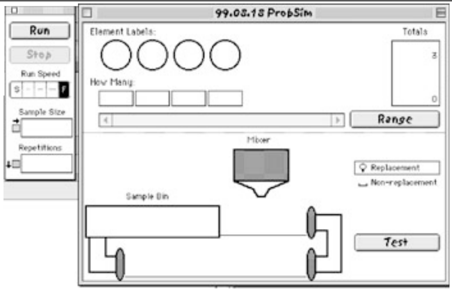


Fig. 1 A task involving the design of a sampling simulation (*left panel*), and the use of Prob Sim (*right panel*) to investigate the issue of whether a particular outcome is unusual

terms of idealized assumptions, involving the identification of a suitable population, a sample, and a random sampling process. This situation, like others we employed in such activities with students, did not describe such aspects and relations. Rather, students had to learn to construe situations in those terms—a process that entailed re-configuring and creatively interpreting information given in the situations. This turned out to be a significant challenge for most students. As illustrated in the following data excerpts, students' progress was tentative and tightly embedded within their interactions with the instructor in the classroom discussions.

4.3 *Constructing Assumptions*

The following illustrative sequence of data excerpts is drawn from classroom discussions that centered on explicating the assumptions for simulating the movie theatre situation. Excerpt 1 illustrates David's sense of overwhelm by the possibilities for assumptions.

Excerpt 1 (Lesson 6):[11]

428. David: I didn't get this question 'cause it, there are so many different things that could happen. Like, what if only half the town goes to see movies? Or uhh what if it's the

[11]The following symbol key is used for discussion protocols: "Instr" denotes an utterance made by the instructor; "[...]" denotes text that is not included in the presented excerpt; "—" denotes that the word or statement immediately preceding it was abruptly halted in speech. Underlined words and statements denote that they were emphasized in speech. Statements enclosed in parentheses generally provide contextual information not captured by utterances, such as participants' nonverbal actions or the duration of pauses or silences in a discussion.

same 2 people every night that he sees? It says he knows 300, but couldn't, like, the same 2 people go see the movie every night?

429. Nicole: Yeah.
430. Instr: Sure, that's right. So that's where you lay—
431. Peter (to David): Good thinking!
432. Instr (continues): You settle all of this in your assumptions. Like, one of the assumptions that you have to make in order to look at this in the abstract, without actually knowing him and the town, is that it's a random process by which the verand—theatre gets filled every week. (3-second pause) Now, it may not in fact be! But in, that's an assumption that you could make that will let you proceed.
433. David: Oh, Ok.
434. Sarah: You also have to assume that he sees everyone that goes to the movie.
435. Instr: Very good! Because if he only sees a small fraction of the people going in, people could be there and he might not see them. (3-second pause) All right. So we're not saying he does, but we're saying in order to proceed we'll make this assumption. Ok, all right. Does that make sense, David?
436. David: Yeah.

Excerpt 1 indicates that David was unable to decide what to assume because he felt lost in a sea of possible choices. David appeared to think of an assumption as a hard fact about the situation, rather than as a working supposition upon which to proceed further. Thus, David's difficulty appeared to be in looking beyond the information given in the situation and reconfiguring it in terms of aspects that are not explicitly given per se, but which are nonetheless necessary to presume. The need to reason hypothetically about a situation as a starting point for designing an investigation of an issue, together with the absence of clear constraints on what could be hypothesized, made the tasks seem too open-ended and ambiguous to some students. The classroom discussions were intended to help students learn to deal with such ambiguity by providing them with opportunities to unpack their implicit assumptions and create new assumptions.

Students' difficulties in deciding what to assume about the given situation were ongoing in these discussions. Decisions were rarely made in a clear-cut manner; instead they often emerged out of relatively arduous negotiations embedded within messy interactions. The subsequent excerpts illustrate this. Excerpt 2 begins with David struggling to make sense of the the italicized part of the central question that was posed as part of the movie theatre situation: "Is it in fact unusual that at least two people Ephram knows attend the movie he shows, *or could people be coming because he is there*?".

Excerpt 2 (Lesson 6):

482. David: Why did you throw in that last part that says "or could people be coming because he is there?" Why did you put that part? That was, that wigged me out (motions with hands above head), I didn't know what to do. It's, like, what is that?
483. Instr: Oh! Well—
484. David: It says (reads) "or could people be coming just because he is there?"
485. Nicole: Yeah! That's my point!
486. Instr: Or for some other reason or another.
487. David: Yeah. I was, like, what is <u>that</u>?
488. Instr: Well, if he always saw—
489. Peter: We have to assume that they're not?

490. Instr (continues): if he always saw 30 people that he knows—a tenth of the people that he knows in this town are there every night, then something's going on, right? (2 second pause). That, I mean—(coughing)

[...]

491. David: Yeah, maybe he's sneaking them in for free?
492. Instr: Perhaps. Something's going on (turns on laptop and window re-appears on screen). Would you expect him to see very many people that he knows? If he knows, if there are 30,000 people and it's a random draw to fill the theatre, would you expect him to see very many people that he knows?
493. Nicole: Well, how many movies are there a night?
494. David: He only knows, like, 1 % of the town, so it's kind of weird that he'd see people, 2 people every single night.
495. Luke: Yeah.
496. Instr: Yeah, he knows 1 % of the town.
497. Instr: Well, we're going to simulate it, we don't know the answer to the question, yet!
498. Kit (to David): I think there's one movie per night.
499. Instr (continues): It might be rare.

The question that students found so problematic in Excerpt 2 was intended to occasion reflection on the reasonableness of the randomness assumption, in the case that the event in question turned out to be statistically unusual. So that, if the event turned out to be unusual, issues could then be raised about whether something other than a random attendance process was at play in the scenario. David and Nicole (lines 482–487) did not, however, interpret things this way. To them this appeared as an isolated question that made little sense and which they could not relate to the greater task. In retrospect, their difficulty may be seen as unsurprising: because the class had not yet investigated whether the event in question might be unusual, these students could not interpret this statement as broaching a possibly relevant issue. The tension that these students experienced drove the instructor to increasingly bring issues of the underlying assumptions out into the discussion.

4.4 Conceiving a Population and a Sample

In the ensuing interaction in Excerpt 2 (line 494), David expressed his "gut level feeling" that the event in question—seeing two or more people at the movie theatre each night—is inconsistent with the given assumption that Ephram knows only 1 % of the 30,000 people in the town. Thus, David's intuition seemed to touch on the idea of drawing a non-representative sample from an underlying population. Moreover, David's intuition also suggests that he was mindful of the 300 acquaintances as a proportion of the entire population, but he did not seem to reason similarly about a sample so as to think of the number of acquaintances as a proportion of 250 people selected. Had he done so and noticed that, say, two or three acquaintances out of 250 is close to 1 %, he might not have found such a result surprising.

Nicole (line 493) also began entertaining possibilities for underlying assumptions, such as "how many movies are we to assume were shown each night?", suggesting that she thought this had important implications for how many acquaintances

Ephram could reasonably expect to see at the movies each night. Thus, Nicole had not yet structured the situation in terms of an unambiguous sampling experiment in which the act of looking into a filled 250-person theatre once per night is *as if* one were recording the composition of a 250-person sample collected from the town's population (assuming random attendance).

In the third episode of the discussion sequence (Excerpt 3), the instructor moved to engage students in choosing values for the Prob Sim parameters in order to simulate the movie theatre situation. The discussion thus turned to making explicit connections between the situation and an idealized sampling experiment, constrained and guided by the structure of the software interface.

Excerpt 3 (Lesson 6):

504. Instr: Now, I'm going to, I'm gonna do this in a way that uhh the guy who wrote this program suggested (sets up Prob Sim to simulate movie theatre scenario). Put a little tiny dot to represent a person in the town who he doesn't know, and a big dot to represent a person in the town that he does know. How many of these dots are there gonna be in this mixer? (points to small dot in left-most element label on the screen displayed in Fig. 2).
505. Luke: 30,000.
506. Instr: No.
507. (Others students chime in unanimously and simultaneously): "27,000".
508. Instr: Yeah, 27,000. No, 29,700.
509. Peter: 29,700.
510. (Instructor enters this value into "how many" slot under first element label)
511. Peter (to others): Mathematical geniuses!
512. (Peter and Nicole laugh in background)
513. Instr: So how many people does—uhh that's because he knows 300 of those 30,000. Right?
514. Luke: Nods. He knows (inaudible) thousand.

[. . .]

515. Instr: Ok, we're gonna take 250 of those people. Right? (assigns this value as sample size in window on screen) Are we taking them with or without replacement?
516. Lesley: Without.
517. Peter: With, with!
518. Luke: No, without replacement.
519. Lesley: With!
520. Nicole: No, you can't, it.
521. Luke: You can't—
522. Nicole: If it's people it has to be—without.
523. Peter: 'Cause they can come back the next night.
524. Instr: No, no, we're talking about one night.
525. Kit (to Peter): Yeah, but not on the same night.
526. Peter: Oh!
527. Luke: Repetitions is (inaudible).
528. Instr: One night. So, is it with replacement or without replacement? (points cursor at "replacement" option in Prob Sim window on screen).
529. Nicole: With.
530. Lesley: With.
531. Luke: Without replacement.
532. Kit: Without.
533. Instr: If a—can a person be in a theatre twice?

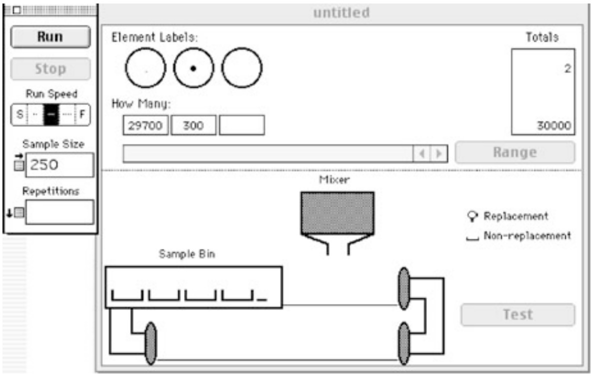


Fig. 2 Prob Sim parameter settings for simulating the movie theatre situation

534. Kit: No.
535. Luke: No.
536. Lesley: No.
537. Instr: Ok, so it's without replacement.
538. Peter: They snuck back in and watched it again.
539. Nicole: Wait.
540. Kit (to Peter): not at the same time you can't.
541. David (to Instructor): You got it on without replacement.
542. Nicole: it's one movie one night.
543. Instr: Yeah.
544. Lesley: I don't understand.
545. Instr: Or, it's just one night. We don't know how many movies, but—
546. Nicole: Well then it's a difference! That's what I asked you.
547. Instr: Ok, then let's say one movie one night. That's a good assumption.

Excerpt 3 began with the instructor using a small dot and a large dot in the software's element labels to represent the population items *acquaintance* and *non-acquaintance*, respectively (Fig. 2). Luke then proposed that the Mixer—the software's metaphor for a population—should contain 30,000 of the small dots (line 505). This suggests that in this context Luke had not conceived of the population as comprising two distinct classes: acquaintances and non-acquaintance. The instructor and other students immediately chimed in with a different answer (lines 506–509). After an estimation error was resolved, the instructor explained to Luke that the population is divided into 300 acquaintances out of 30,000 total people, to justify choosing 29,700 as the number of small dots in the mixer.

The discussion then turned to assigning values to the sampling parameters. The instructor proposed that sample size should be 250. There were no objections to this, and the issue then became whether to sample with or without replacement (line 515). Here, students held different opinions, but direct evidence of *how* they were thinking about the issue is limited.[12] It seems, however, that a source of the differences might be attributed to students holding different assumptions about how

[12]The distinction between sampling with and without replacement had been established in classroom discussions prior to this point in the teaching experiment.

many movies are shown each night. In line 523, Peter appeared to assume that people could return to the movie theatre on different nights; an assumption suggesting that he had not structured the situation in terms of what might occur on an individual night as a distinct unit. Rather, Peter seemed to be considering events that could occur across several nights. Peter's utterance in line 538 indicates that even when restricting himself to considering an individual night, he was thinking of contingencies that suggest he was unclear on *what* constituted a sample in the scenario. "Could one sample be like one full audience, or should it be thought of as all audiences in one night?" We speculate that such underlying questions may have been the kind Peter was contemplating without resolution.

Peter's difficulty may be interpreted as one of making an idealized assumption about the situation. Reformulating a situation in terms of idealized assumptions is a hallmark of the mathematical modeling process. This entails choosing to ignore certain contingencies that, while perhaps realistic or plausible, may make the situation too unwieldy to model. One has to consciously make discerning choices about how to appropriately simplify a situation so as to make it amenable to model. Peter's comments presumably reflect an underlying difficulty in making such judicious choices. Evidently for Peter, neither the constraints of Prob Sim nor the instructor's scaffolding efforts were sufficient to help him structure the situation as an unambiguous sampling experiment in which a sample consisted of a 250-person audience that attended a single movie in one night. The metaphor of population as a mixer may have been relatively transparent, but conceptualizing the situation in terms of a clear-cut sample, the selection of which he might imagine repeating, was problematic.

Although the best available evidence in Excerpt 3 is of Peter's thinking, the other students' utterances collectively indicate that the group was generally indecisive about whether to sample with or without replacement. Students such as Lesley and Nicole, who flitted from one sampling option to the other, were evidently unsettled on what to take as a sample in the scenario; their assumptions were still formative and highly unstable. Eventually, the instructor and Nicole negotiated the assumption that a sample should be a single movie in a single night (lines 542–547)—a simplification consistent with sampling without replacement.

The data excerpts presented here are representative of the extended classroom discussions that ensued around other simulation-design tasks we employed in an effort to provoke students to re-construe situations as stochastic experiments. These excerpts illustrate the messy and somewhat disorderly nature of classroom conversations marked by uneven participation by students relative to each other and to the instructor. Despite this, collectively, these excerpts illustrate how construing a contextual situation as a statistical experiment can be a highly non-trivial activity for students. Even under conditions of heavy instructional guidance, entailing a supportive classroom environment in which it was normative to dissect ideas, and the use of software intended to structure and constrain the activity in productive ways, students experienced considerable difficulties. These difficulties extended to several aspects and various levels of detail of situations, but particularly to making discriminating decisions about underlying assumptions. The connections between a given

Table 1 Frequency of student responses to the question "Does this describe a probabilistic situation?"

Response	Frequency
Yes	5
No	3

situation *as it is described* and its construal as a stochastic experiment were generally not transparent for students.

4.5 A Subsequent Attempt to Address Students' Difficulties

As part of our effort to address these difficulties in the second teaching experiment, our instruction first focused on conducting whole-class discussions aimed at having students ascertain whether given questions or situations were "probabilistic" per se. By "probabilistic" we mean that a situation involves a stochastic experiment that produces an identifiable outcome whose long-run relative frequency could be determined by repeating an experiment under essentially identical conditions. For each question or situation presented, the students' task was to justify their decision as to whether the situation was probabilistic. Moreover, we had students restate each non-probabilistic question or situation so as to transform it into a probabilistic one, and to explain why in some cases it was not possible to so transform it. Two examples of such questions are:

Question 1: *What is the probability that on a toss of three coins you will get two or more heads?*
Question 2: *What is the probability that the next U.S. Attorney General is a woman?*

We engaged students in construing situations in probabilistic terms over four consecutive lessons before having them return to using Prob Sim as a simulation tool. These reconstrual activities were interspersed with computer demonstrations of the law of large numbers, showing graphical depictions of the cumulative relative frequency of a sampling outcome of interest stabilizing to a particular value over a large number of trials of a random experiment (e.g., selecting samples of balls from a population of red and blue balls, the proportion of which was to be inferred). Due to space constraints, we only illustrate aspects of students' thinking around these reconstrual activities with a few excerpts from individual interviews conducted with students at the end of that phase of the teaching experiment. In the interviews, we asked students whether Question 1 (above) describes a probabilistic situation. In fact, this question does not describe a probabilistic situation per se; it refers to a single trial of a 3-coin toss, and makes no mention of repeating it or what fraction of the time one might expect to observe two or more heads in a large number of (identical) repetitions. Table 1 displays the frequency of students' responses.

Five of the students thought the question described a probabilistic situation, and three thought it did not. However, as revealed by their justifications, students were surprisingly (to us) flexible in their thinking. All of those who responded “No” thought they were being asked about the question *as it is stated* and explained that it mentioned only one toss or trial. Moreover, with little or no further prompting, most of these students were easily able to restate the question in a manner that involved repeating the 3-coin toss experiment a large number of times and finding the fraction of the time with which it came up two or more heads. A representative example of such a response is shown in the brief interview excerpt below:[13]

Excerpt 4

Peter: (reads) What is the probability that on a toss of three coins you will get two or more heads?
Int: Yeah. Now, what kind of situation –is that a, is that a probabilistic situation?
Peter: N—no.
Int: Okay, w—why, why not?
Peter: Because—whatever, I guess it’s just *zero* or one. You do or you don’t.
Int: Okay. You do or you don’t?
Peter: Get two or more heads.
Int: Okay.
Peter: Just on that, that toss of three coins. I guess, to make it a probabilistic situation, you have to toss those three coins a large number of times.
Int: Okay.
Peter: And then see how many times you got two or more heads.
Int: All right and, and then—
Peter: And then see what fraction of the time—
Int: That you get two or more heads?
Peter: Yeah.

All of the students who responded “Yes” *interpreted* Question 1 as if it involved a probabilistic situation, explaining that one can repeat the coin toss experiment a large number of times and find the fraction of times that it comes up with two or more heads. An illustrative example of such a response is shown in the interview excerpt below. The second part of the excerpt also indicates that this particular student was clearly able to recognize that the situation *as stated* is not probabilistic and why that is so:

Excerpt 5

Sarah: (reads) What is the probability that on a toss of three coins, you will get two or more heads?
Int: Now, uhm, does that seem to be a probabilistic situation—the way that it is described there—to you?
Sarah: Uhm, yes.
Int: How, how so?
Sarah: Because you can, uh, experiment, uh, and you can—you can experiment and you’ll get, like, an outcome that will show, uh, what fraction of the time, uhm, you get two or more heads.

[…]

Int: So, what you just described there is not what is actually written on—in—this statement, right?

[13]“Int” denotes an utterance made by the interviewer.

Sarah: Right.
Int: Okay. So you—*you've*, you've thought about, about this in probabilistic terms.
Sarah: Uhm huh (affirmative).
Int: But the way it is actually described–
Sarah: It's not.
Int: —it's not. And why is it not?
Sarah: Because–
Int: What about this statement makes it *non*-probabilistic?
Sarah: Because, uhm, you know, it's only something happening once, and you can only get one, one outcome or the other.
Int: What is the thing that's happening once, here?
Sarah: You toss three coins.
Int: Okay.
Sarah: You can only do that once and either you'll get two heads or you won't.

These results are encouraging. Our instructional effort to focus students' attention more narrowly on thinking about situations per se, and reconstruing the situations in stochastic terms, without the added conceptual burden of having to design a simulation of them appears to have been productive. A possible related factor in their success here is that more of the situations we presented to them in the first phase of the second experiment were of a form that may have cued their prior ideas or experiences about canonical stochastic situations (e.g., tossing fair coins or dice). This is in contrast to the arguably more complex and ambiguous situations they encountered in the first experiment (e.g., the Movie Theatre situation).

A next stage in our research will be to understand how robust were students' stochastic conceptions evidenced at this early stage of the second experiment, and the extent to which those conceptions may or may not have supported students' efforts to deal with increasingly complex situations and probability tasks as the experiment progressed into its later phases.[14]

5 Conclusion

We began this chapter with a review of two distinctive clusters of research on probability and probabilistic reasoning. In doing so, we hoped to elucidate conceptual, psychological, and instructional challenges we face in teaching and learning probability. We first surveyed a selection of the mathematical theories of probability. Our survey provided a background for making sense of the psychological studies in probabilistic reasoning and for clarifying the central issues of designing learning objectives in probability instruction. We concurred with a number of researchers' thesis (Konold 1989; Shaughnessy 1992) that when designing instruction probabilistic reasoning is better thought of as creating models of phenomena.

[14]Our second teaching experiment consisted of 22 lessons, and eventually addressed the concepts of conditional probability, independence, and Bayes Theorem.

Next, we reviewed literature on epistemological and psychological studies in probabilistic reasoning. We summarized a number of factors considered to be pertinent in students' development of probabilistic reasoning. This included developmental constraints, world view and orientations (towards uncertain situations), beliefs of what probability is and is about, intuitions and judgment-heuristics, and proportionality. We proposed that a key conceptual scheme in developing coherent probabilistic reasoning is that of a stochastic conception: a conception of probability that is built on the concepts of random process and distribution.

In the last section, we presented evidence of a group of high school students' thinking that emerged in relation to their engagement in classroom teaching experiments designed to promote their developing the scheme of ideas that comprise a stochastic conception of events and probability. This evidence highlighted challenges in learning (and teaching) those ideas centered around developing the ability to construe situations as stochastic experiments. Our first experiment used the activity of designing sampling simulations for the Prob Sim micro-world as a context within engaging students with the intricacies of construing ambiguous situations as stochastic experiments. These intricacies included making idealized assumptions, construing an appropriate population, and determining a sample and sampling method, all of which were considerably challenging for students. Our second experiment eased the difficulty somewhat by having students practice reconstruing given situations in probabilistic terms, without requiring the design of simulations of them. Although we are encouraged by the evidence that all students were able to do this, the robustness and carrying power of this ability for understanding and dealing coherently with more complicated situations and probability problems is as of yet unknown.

An overarching implication that we draw from our review of the literatures surveyed here, and that is reflected in the instructional efforts in our teaching experiments, is that probability instruction should emphasize students' imagery, conceptions, and beliefs when they learn probability. Supporting students' development of coherent probabilistic reasoning entails conceiving of probability as ways of thinking, instead of as a set of disconnected concepts and skills. As a result, instructional programs should deal directly with students' epistemological and conceptual obstacles in learning. The design of such programs will require an understanding of how instruction helps students overcome these obstacles, and the challenges they may face as they engage with the instruction. We view the instructional efforts reported here, the evidence of our students' challenges, and our future research involving classroom teaching experiments in probability learning, as part of an endeavor to contribute to such understanding.

References

Biehler, R. (1994). *Probabilistic thinking, statistical reasoning, and the search for causes—do we need a probabilistic revolution after we have taught data analysis?* Paper presented at the 4th international conference on teaching statistics, July 2004, Minneapolis, Minnesota.

David, F. N. (1962). *Games, gods, and gambling: the origin and history of probability and statistical ideas from the earliest times to the Newtonian era*. New York: Hafner.
de Finetti, B. (1970). *Theory of probability: a critical introductory treatment*. London: Wiley.
Doob, J. L. (1996). The development of rigor in mathematical probability. *The American Mathematical Monthly*, *3*(7), 586–595.
Falk, R., & Konold, C. (1992). The psychology of learning probability. In F. S. Gordon & S. P. Gordon (Eds.), *Statistics for the twenty-first century* (pp. 151–164). Washington: Math. Assoc. of America.
Fine, T. L. (1973). *Theories of probability: an examination of foundations*. New York: Academic Press.
Fischbein, E. (1975). *The intuitive sources of probabilistic thinking in children*. Dordrecht: Reidel.
Fischbein, E., & Gazit, A. (1984). Does the teaching of probability improve probabilistic intuitions? *Educational Studies in Mathematics*, *15*, 1–24.
Fischbein, E., Nello, M. S., & Marino, M. S. (1991). Factors affecting probabilistic judgments in children and adolescents. *Educational Studies in Mathematics*, *22*, 523–549.
Fischbein, E., & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for Research in Mathematics Education*, *28*, 96–105.
Garfield, J., & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: implications for research. *Journal for Research in Mathematics Education*, *19*, 44–63.
Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is relevant for psychology (and vice versa). In G. Wright & P. Ayton (Eds.), *Subjective probability*. New York: Wiley.
Gigerenzer, G. (1996). The psychology of good judgment: frequency formats and simple algorithms. *Journal of Medical Decision Making*, *16*, 273–280.
Gigerenzer, G. (1998). Ecological intelligence: an adaptation for frequencies. In D. D. Cummins & C. Allen (Eds.), *The evolution of mind*. Oxford: Oxford University Press.
Gillies, D. (2000). *Philosophical theories of probability*. New York: Routledge.
Good, I. J. (1965). *The estimation of probabilities: an essay on modern Bayesian methods*. Cambridge: MIT Press.
Gould, S. J. (1992). *Bully for brontosaurus: further reflections in natural history*. New York: Penguin.
Green, D. (1979). The chance and probability concepts project. *Teaching Statistics*, *1*(3), 66–71.
Green, D. (1983). School pupils' probability concepts. *Teaching Statistics*, *5*(2), 34–42.
Green, D. (1987). Probability concepts: putting research into practice. *Teaching Statistics*, *9*(1), 8–14.
Green, D. (1989). School pupils' understanding of randomness. In R. Morris (Ed.), *Studies of mathematics education: the teaching of statistics*, Paris, UNESCO (pp. 27–39).
Hacking, I. (1975). *The emergence of probability*. Cambridge: Cambridge University Press.
Hacking, I. (2001). *An introduction to probability and inductive logic*. Cambridge: Cambridge University Press.
Harel, G., & Confrey, J. (1994). *The development of multiplicative reasoning in the learning of mathematics*. Albany: SUNY Press.
Hawkins, A., & Kapadia, R. (1984). Children's conceptions of probability: a psychological and pedagogical review. *Educational Studies in Mathematics*, *15*, 349–377.
Hendricks, V. F., Pedersen, S. A., & Jorgensen, K. F. (2001). *Probability theory: philosophy, recent history and relations to science*. Dordrecht: Kluwer Academic.
Hertwig, R., & Gigerenzer, G. (1999). The "conjunction fallacy" revisited: how intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making*, *12*, 275–305.
Kahneman, D., & Tversky, A. (1972). Subjective probability: a judgment of representativeness. *Cognitive Psychology*, *3*, 430–454.
Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*, 237–251.
Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, *47*, 263–291.

Kahneman, D., & Tversky, A. (1982). Evidential impact of base rates. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: heuristics and biases* (pp. 153–160). Cambridge: Cambridge University Press.

Kolmogorov, A. N. (1950). *Foundations of the theory of probability*. New York: Chelsea.

Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction*, *6*(1), 59–98.

Konold, C. (1995). Issues in assessing conceptual understanding in probability and statistics. *Journal of Statistics Education*, *3*(1).

Konold, C. (2002). *Prob SIM user guide*. Amherst: Author.

Konold, C., & Miller, C. (1996). *Prob Sim. Computer program*, Amherst, MA.

Konold, C., Pollatsek, A., Well, A., Lohmeier, J., & Lipson, A. (1993). Inconsistencies in students' reasoning about probability. *Journal for Research in Mathematics Education*, *24*(5), 393–414.

Krüger, L. (1987). *The probabilistic revolution*. Cambridge: MIT Press.

Liu, Y., & Thompson, P. W. (2007). Teachers' understandings of probability. *Cognition and Instruction*, *25*(2), 113–160.

Metz, K. (1998). Emergent understanding and attribution of randomness: comparative analysis of the reasoning of primary grade children and undergraduates. *Cognition and Instruction*, *16*(3), 285–365.

Moore, D. (1990). Uncertainty. In L. A. Steen (Ed.), *On the shoulders of giants: new approaches to numeracy* (pp. 95–137). Washington: National Academy Press.

Nilsson, P. (2003). Experimentation as a tool for discovering mathematical concepts of probability. In *Proceedings of the 3rd conference of the European Society for Research in Mathematics Education*, Bellaria, Italy.

Piaget, J., & Inhelder, B. (1951/1975). *The origin of the idea of chance in children*. London: Routledge & Kagan Paul.

Piatelli-Palmarini, M. (1994). *Inevitable illusions: how mistakes of reason rule our minds*. New York: Wiley.

Piccinato, L. (1986). De Finetti's logic of uncertainty and its impact on statistical thinking and practice. In P. K. Goel & A. Zellner (Eds.), *Bayesian Inference and decision techniques: essays in honor of Bruno de Finetti*. New York: Elsevier.

Ritson, I. L. (1998). *The development of primary school children's understanding of probability*. Unpublished doctoral dissertation, Queen's University, Belfast.

Saldanha, L. A. (2004). *"Is this sample unusual?": an investigation of students exploring connections between sampling distributions and statistical inference*. Unpublished doctoral dissertation, Vanderbilt University, Nashville, TN.

Saldanha, L. (2010). Conceptual issues in quantifying expectation: insights from students' experiences in designing sampling simulations. In *Proceedings of the 8th international conference on teaching statistics (ICOTS-8)*, Ljubljana, Slovenia, July 2010.

Saldanha, L. (2011). Aspects of conceiving stochastic experiments. In B. Ubuz (Ed.), *Proceedings of the 35th conference of the international group for the psychology of mathematics education* (Vol. 4, pp. 97–104). Ankara: PME.

Saldanha, L., & Thompson, P. W. (2002). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics*, *51*(3), 257–270.

Saldanha, L., & Thompson, P. (2007). Exploring connections between sampling distributions and statistical inference: an analysis of students' engagement and thinking in the context of instruction involving repeated sampling. *International Electronic Journal of Mathematics Education*, *2*(3), 270–297.

Schwartz, D. L., & Goldman, S. R. (1996). Why people are not like marbles in an urn: an effect of context on statistical reasoning. *Applied Cognitive Psychology*, *10*, 99–112.

Schwartz, D. L., Goldman, S. R., Vye, N. J., & Barron, B. J. (1998). Aligning everyday and mathematical reasoning: the case of sampling assumptions. In S. P. Lojoie (Ed.), *Reflections on statistics: learning, teaching, and assessment in grades K-12* (pp. 233–274). Mahwah: Erlbaum.

Sedlmeier, P. (1999). *Improving statistical reasoning: theoretical models and practical implications*. Mahwah: Erlbaum.

Shaughnessy, J. M. (1977). Misconceptions of probability: an experiment with a small-group, activity-based, model-building approach to introductory probability at the college level. *Educational Studies in Mathematics*, *8*, 295–316.

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.

Todhunter, I. (1949). *A history of mathematical theory of probability: from the time of Pascal to that of Laplace*. New York: Chelsea.

von Mises, R. (1957). *Probability, statistics, and truth*. London: Allen & Unwin.

von Plato, J. (1994). *Creating modern probability: its mathematics, physics, and philosophy in historical perspective*. Cambridge: Cambridge University Press.

"It Is Very, Very Random Because It Doesn't Happen Very Often": Examining Learners' Discourse on Randomness

Simin Jolfaee, Rina Zazkis, and Nathalie Sinclair

Abstract We provide an overview of how the notion of randomness is treated in mathematics and in mathematics education research. We then report on two studies that investigated students' perceptions of random situations. In the first study, we analyze responses of prospective secondary school teachers who were asked to provide examples of random situations. In the second study, we focus in depth on participants' perceptions of randomness in a clinical interview setting. Particular attention is given to the participants' ways of communicating the idea of randomness, as featured in the gestures that accompanied their discourse. We conclude that particular consideration of the notion of randomness—as intended in statistics and probability versus everyday uses of the term—deserves attention of instructors and instructional materials.

Keywords Randomness · Uncertainty · Probability · Gestures · Example space · Definition · Prospective secondary school teachers · Learner generated examples

1 Introduction

The notion of randomness is central to the study of probability and statistics and it presents a challenge to students of all ages. However, it is usually not defined in textbooks and curriculum documents, as if the meaning of randomness should be captured intuitively. In fact, the word "random" does get used in everyday language, but not always in the same way that it is used in mathematics. Even in mathematics, the notion of randomness has been a challenging one—not only did it emerge relatively recently in the history of mathematics, it has also undergone various attempts to be adequately defined. Given the importance of randomness in the study of probability and its complexity as a concept, our goal in this chapter is to better understand the ways learners use and talk about it.

In this chapter, we first provide an overview of some of the ways in which randomness is defined in mathematics—these aspects of randomness will help struc-

S. Jolfaee (✉) · R. Zazkis · N. Sinclair
Simon Fraser University, Burnaby, Canada
e-mail: schavosh@sfu.ca

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_21,

ture our analysis of learners' uses and descriptions of it. We then provide a brief overview of the research in mathematics education and highlight the main resources that learners use to explain randomness. Following this, we present two empirical studies involving prospective teachers and undergraduate students, each aimed at further probing understandings of randomness using different methodological approaches.

2 Randomness in Mathematics

Randomness (and probability theory more generally) is a human construct created to deal with unexplained variation. We use probability theory to model and describe phenomena in the world for which there is a lack of deterministic knowledge of the situation, assuming that they had been randomly generated. Thus, what probability is can only be explained by randomness, and what randomness is can only be modeled by means of probability. The notion of probability is heavily based on the concept of a random event and statistical inference is based on the distribution of random samples. Often we assume that the concept of randomness is obvious but, in fact, even today, experts hold distinctly different views of it.

The study of random sequences was revived in the field of mathematics when it became clear that new ideas from set theory and computer programming could be used to characterize the complexity of sequences. Von Mises in 1952 based his study of random sequence on the intuitive idea that a sequence is considered to be random if we are convinced of the impossibility of forecasting the sequence in order to win in a game of chance. This notion of randomness (called by von Mises the *impossibility of a gambling system*) is closely and fundamentally tied to the notion of independence. On the other hand, the inability to gamble successfully encapsulates an intuitively-desirable property of a random sequence: its unpredictability.

Later, in 1966, von Mises suggested that a random sequence does not exhibit any exceptional regularity effectively testable by any possible statistical test. This approach is similar to Kolmogorov and Chaitin's (Li and Vitany 2008) vision of a random sequence as a highly irregular or complex sequence (also called algorithmically incompressible or irreducible) that the sequence cannot be reproduced from a set of instructions that is shorter than the sequence itself. For example, the following sequence of 1s and 0s—1010101010101010101010...—is not random because there exists a short description for it: write down 10 infinitely many times. It is noteworthy that this definition relies heavily on language (ordinary or computer-based) and testing methods.

Durand et al. (2003) found it rather surprising that algorithms are involved in defining random sequences, since probability theory does not use the notion of algorithm at all, and proposed more general definitions based only on set theory. One of the nuances their work brings into the definition of randomness is the notion of typicalness according to which a random sequence is a typical representative of the class of all sequences. In order to be a typical representative, it is required to have no specific features distinguishing the sequence from the general population of the sequences.

In contemporary mathematics, the word 'random' is mostly used to describe the output of unpredictable physical processes (Wolfram Math World). These physical processes are very familiar and include flipping coins and throwing dice. The randomness that is obtained from computer-generation is often called pseudorandom, in part because someone knows the rules of the system that produces the numbers. As such, the very notion of random is intimately tied to the machines that are used to produce individual events (like the result of a flipped coin or the digit between 0 and 9). In addition to unpredictability, the term 'random' usually means random with respect to a uniform distribution. Other distributions are also possible, and this relation of randomness to a particular distribution is what enables the theoretical probabilities that can predict the outcomes of a large number of random events. Without the notion of distribution, one is left simply with unpredictability, which can be found in many everyday situations (like the weather).

In contrast to these more complex definitions, Bennett (1998) proposed a more practical definition of randomness (of a sequence): a sequence is random either by virtue of how many and which statistical tests it satisfies (of which there are many) or by the virtue of the length of the algorithm necessary to describe it (this latter is known as Chaitin–Kolmogorov's complexity measure (Li and Vitany 2008)). According to Wolfram (2002), this latter characterization (the incompressibility of data) is the most valid definition, showing remarkable consistency. It has also been proven that this definition covers a large number of the known statistical tests in effect. By the early 1990s, it had thus become accepted as the appropriate definition of randomness (of a sequence).

One drawback of these definitions is that they emphasize process (the machines and/or tests that are used to create/test) as opposed to structure. In contrast, in *Structure and Randomness*, the mathematician Terrence Tao (2008) describes sets of objects as random if there is no recursiveness of the information between these set of objects. He compares the notion of pseudorandomness with that of structure to distinguish between sets of objects. Structured objects, on the one hand, are those with a high degree of predictability and algebraic structure, for example, the set $A = \{\dots, -5, -3, -1, 1, 3, 5, \dots\}$ is structured since if some integer n is known to belong to A, one knows a lot about whether $n + 1$, $n - 1$, etc. will also belong to A. On the other hand, a set of pseudorandom objects is highly unpredictable and lacks any algebraic structure. For example, consider flipping a coin for each integer number and defining a set B to be the set of integers for which the coin flip results in heads; in such a set B, no element conveys information whatsoever about the relation of $n+1$, $n+2$, etc. with respect to B. While this definition of randomness does not explicitly mention an underlying distribution, it does underscore the relation (or lack thereof) elements of a set have with each other.

Coming from a mathematical modeling point of view, Tsonis (2008) has studied and produced many different examples of randomness (trying to create random sequences is an important part of trying to understand and describe randomness in mathematics). He has proposed three different types of randomness: (i) the first type arises from not knowing the rules of a system (whether the rules are inaccessible or non-existing); (ii) the second type arises from a chaotic system, a system which is sensitive to the initial conditions; (iii) the third type is known as a stochastic process, where there is some kind of external environmental component whose essentially uncountable agents (external noise) continually affect the system with their actions. These categories of randomness are interesting in that they relate more closely to everyday understandings of randomness, as we demonstrate below.

There is no agreed-upon definition of randomness that can provide necessary and sufficient conditions for all the kinds of randomness in finite or infinite sequences and sets. Even when considering the most famous examples of random sequences of numbers, namely decimal digits of π, we know that its first 30 million digits are very uniformly distributed and its first billion digits pass the "diehard tests" (an old standard for testing random number generators, see Marsaglia 2005), which means that the sequence is random in the same sense that the outcome of a fair die is random (known range, unpredictability of the outcomes and uniform distribution). However, the same sequence of numbers doesn't qualify as random according to Kolmogorov–Chaitin's criteria, since there exists a method that describes the sequence in a much shorter length than the sequence itself, Bailey et al. (1997a, 1997b). Despite these definition issues, the educational research suggests that people's difficulty with randomness lies mainly in their inability to generate and/or recognize examples of randomness. Of course, if they knew the definition, they might not have this difficulty. But, as we show in the next section, people often talk about random behaviors and properties in ways that diverge from the mathematically-accepted discourse.

3 Randomness in Mathematics Education

The psychologists Kahneman and Tversky (1972) discussed perceptions of randomness with respect to sequences of coin flips. They suggested that people's intuitive notions of randomness are characterized by two general properties: irregularity (absence of systematic patterns in the order of outcomes as well as in their distribution) and local representativeness (similarity of a sample to the population). They pointed out that local representativeness is a belief that the law of large numbers applies to small numbers as well and wrote that this belief "underlies erroneous intuitions about randomness which are manifest in a wide variety of contexts" (p. 36).

According to Kahneman and Tversky's own description, "random-appearing sequences are those whose verbal description is longest" (p. 38). By this, the authors refer to the fact that in dictating a long sequence of outcomes one would necessarily use some shortcuts, such as "4 Heads" or "repeat 3 times Tails–Heads." However, short runs and frequent switches—that characterize apparent randomness—minimize the opportunities for shortcuts in verbal descriptions. Therefore, apparent randomness was seen as "a form of complexity of structure" (ibid.). This is consistent with Kolmogorov and Chaitin's (Li and Vitany 2008; Vitanyi 1994) view described above.

Researchers in mathematics education continued the tradition, started by psychologists, of using sequences of binary outcomes to make inferences about people's conceptions of randomness. For example, Falk and Konold (1998) investigated the subjective aspects of randomness through tasks in which people were asked to simulate a series of outcomes of a typical random process such as tossing a coin (known as a generating task) or rating the degree of randomness of several sequences (known as perception tasks). Their findings indicate that perceived randomness has several subjective aspects that are more directly reflected in perception tasks "since people might find it difficult to express in generation what they can recognize in perception" (p. 653).

In continuing the investigation of perception tasks, Battanero and Serrano (1999) studied perceptions of randomness of secondary school students, ages 14 and 17, $n = 277$. They presented students with eight items, four involving random sequences and four involving random 2-dimensional distributions. Specifically, the students were shown four sequences of Heads and Tails of length 40, and asked to guess which were made up and which were generated by actual toss of a coin. Further, the participants were presented with a 2-dimensional 4×4 grid, in which results of the "counters" game were recorded (this game involves placing 16 counters numbered 1–16 in a bag, choosing a counter, marking its number in a corresponding cell of the grid and returning the counter to the bag, then repeating 16 times). The participants were asked which grids were the result of an actual game and which, in their opinion, were made up. Their arguments for deciding which results appeared random were recorded and analyzed. Unpredictability and irregularity were the main arguments in support of randomness. This research reveals complexities in the meaning of randomness and it also shows that some students' arguments are in accord with the interpretations attributed to randomness throughout history, such as lack of known cause.

Using computer-based activities, Pratt and Noss (2002) observed 10–11 year-old children using four separable resources for articulating randomness: unsteerability, irregularity, unpredictability, and fairness. While the last three components were mentioned above, unsteerability, the first component, was described as personal inability to influence the outcomes. Further, children often combined unpredictability with unsteerability, where "unpredictability was usually seen as the outcome of uncontrolled input" (p. 464). Moreover, children used these resources interchangeably where different settings initially triggered different resources. Pratt and Noss suggested that intuitive notions of randomness had features analogous to diSessa's (1993) corresponding phenomenological primitives (that is, "multitudinous, small pieces of knowledge that are self-evident, not needing justification and weakly connected" (p. 466)).

Researchers appear to agree that the setting and context significantly influences learners' decision making in probabilistic situations (e.g., Chernoff and Zazkis 2010; Fischbein and Schnarch 1997). However, the notion of randomness was investigated in studies where the context was constructed by researchers. What happens when learners produce their own contexts? In order to answer this question, we designed two empirical studies in which we could analyze learners' perceptions of randomness where the context is not predetermined.

4 Learners' Discourse on Randomness

In order to study learners' discourse on randomness, we conducted two separate empirical studies, each based on different aspects of mathematical knowing. The first draws on Watson and Mason's (2005) suggestions that knowing a particular mathematical concept involves being able to generate a wide variety of examples of the concept, rather than, or in addition to, the ability to provide a definition or description of the concept. Zazkis and Leikin (2008) suggest that learner-generated examples serve not only as a pedagogical tool, as advocated by Watson and Mason (2005), but also as an appropriate research lens to investigate participants' conceptions of mathematical notions. Thus, our first research question attempts to find out how learners exemplify random phenomena and what features of randomness are present in their examples.

Our second study draws on the recent theories of multimodality, which assert that mathematical thinking involves the use of a wide range of modes of expression, including language but also gestures, diagrams, tone of voice, etc. (Arzarello et al. 2009). Given the important connections between gestures and abstract thinking, we are particularly interested in the kinds of gestures learners use to express their ideas of randomness. Based on Sinclair and Gol Tabaghi's (2010) study of mathematicians' use of gesture, we anticipate that learners' gestures might communicate some of the temporal and imagistic aspects of randomness that can be difficult to express in words. Thus, our second research question seeks to find out how learners communicate ideas of randomness through a broad multimodal discourse.

Study 1: Exemplifying Randomness In this section, we consider examples of randomness generated by prospective secondary school mathematics teachers. From 40 participants, we collected 38 different examples. We note with some surprise that only six examples had mathematical context, that is, were related on situation that can be seen as conventional mathematical in the discussion of probability. By "mathematical context" we refer to examples in which the theoretical probability of the exemplified random event can be calculated, such as:

- When playing with a standard pack of 52 playing cards, what is the probability of getting a "King of Hearts" if I choose a card at random?

Though there was no specific request that examples had to be related to a mathematical context, we find the paucity of such examples rather surprising given that the data was collected from the population of mathematics teachers in a mathematics-related course. Our further surprise is that most of the mathematical examples (5 out of 6) related to card games or gambling. We speculate that this is because the participants understood or experienced the effects of such randomness on the desired outcome.

In what follows, we examine the examples generated by participants first using Tsonis' (2008) tripartite model and then by using the characteristics of randomness described by Pratt and Noss (2002). We then discuss features evident in our examples that differ from those in the literature.

Types of Randomness According to Tsonis According to Tsonis, randomness of the first type arises from not knowing the rules of a system either in the case of non-existence of such rules, or of the inability to access such rules because of irreversible programs or procedures that have inhibited us from getting access to those rules. The following are illustrative examples of this kind of randomness (from our data).

- Babies being born with autism
- It will rain tomorrow

While there are scientific explanations for the causes of these phenomena (autism, rain), not knowing the explanations resulted in perception of randomness.

Randomness of the second type arises from a chaotic system, a system which remembers its initial conditions and the evolution of the system is dependent on those; in other words, the system is very sensitive to the initial conditions: only if we had infinite precision and infinite power, we would be able to predict such systems accurately. "Rolling snake eyes on a craps table" (an example from our data) is a perfect fit for this category since from a very precise mathematical standpoint it is possible to propose a mechanical model that describe a fair die and build its motion equations and take into account reasonably all of the parameters that matter such as the speed and angle at the tossing point (with a finite accuracy), the viscosity of air, the gravity at the place of experiment, the mass distribution of the coin, the elasticity factors of the table (or the ground at which the coin hits) and eventually show that the outcome of the two dice is uniquely predictable (Strzalko et al. 2009). At the same time rolling two dice several times is completely random due to the impossibility of controlling initial conditions when rolled by human, which makes the randomness inherent to dice rolling examples the randomness of second type. On a less pedantic level, this category can include examples in which the unpredictability of the situation is closely related to the idea that small changes in the initial conditions can produce big differences in the outcome. For example:

- A meteor landing on your house
- An earthquake occurring in Vancouver

We find in examples of this kind the perception of a chaotic system, where the outcome is unpredictable.

Randomness of the third type is in which there is some kind of external environmental component whose essentially uncountable agents (external noise) continually affect the system with their actions. For instance, consider the time a specific person (Sarah) spends in a shopping mall each time: We reasonably understand how the shopping enterprise takes place so it is not randomness of the first type, there are no initial conditions that the system is super sensitive to, therefore it is not randomness of the second type, the thing is that Sarah's shopping time depends on several environmental (the number is not known) factors such as how many stores are having sales, what are the line ups like in the stores that she shops, and whether she runs into a friend, etc. Either of these factors could be considered as simple and rather deterministic systems but the collective behavior of many interacting systems may be very complicated. In terms of our data, this category includes examples in which the main causes and rules are known but the variability or unpredictability arises from the exposure of those rules to external agents. For example:

- Someone will miss our next class
- Leaves falling off the tree when it is not autumn

A Note on Fitting Examples into Tsonis' Model When trying to categorize examples provided by the participants by the types of randomness in Tsonis' model, one important aspect of probabilistic thinking drew our attention: different people with different backgrounds and knowledge find things random at different levels for different reasons. Consider the "babies born with autism" example. At one level, this may illustrate randomness of the first type, since the causes of autism (the rules of the system) and why some babies are born with or without it are unknown to a person with no background in biology. At another level, for a person who is aware that autism has a strong genetic basis, and can be explained by rare combinations of common genetic variants, randomness of the second type seems a better fit. Yet again, the same example could be considered as randomness of the third type for someone who knows more about the technicalities of mutation (that mutation could be caused by external agents such as radiation, viruses and some chemicals). Given the ambiguity involved in this categorization, we considered what could be gained by attending to different attributes of randomness.

Features of Randomness According to Pratt and Noss The resource of fairness identified by Pratt and Noss (2002) was the only one not present in our data. However, we provide examples of each of the other three attributes:

Unsteerability

- It is sunny in Ottawa on Canada Day
- It will rain tomorrow
- A tsunami being triggered by an earthquake

Some examples in this category related to weather conditions and involved the inability to influence the outcome.

Irregularity

- A particular bus will stall or break down in any given hour of service
- Someone will miss our next class
- Getting caught talking/texting on your cell phone while driving
- Giving flowers to your wife when it is not Valentine's, birthday, or an anniversary

The feature of regularity is usually described in terms of sequence outcomes that have no apparent pattern, rather than in terms of a single event. However, in these examples, the sequencing is implied by a routine, such as taking a bus to school daily, attending the class, texting while driving and not getting caught, acknowledging certain dates with flowers. Here irregularity can be seen as a deviation from the routine.

Unpredictability

– Dying in an airplane accident
– Being struck by lightning
– Getting a royal flush in poker

A large number of examples attended to events that are unpredictable, with a frequent reference to accidents or natural disasters.

Though we attempted to choose examples that best illustrate each of the features, it is important to note possible overlaps. The event of a particular bus breaking down is unsteerable (by the participants) as well as irregular and unpredictable. The same consideration applies to getting a royal flush in the game of poker.

A notable feature of the generated examples, which does not appear in the definitions of mathematicians or in the mathematics education literature, is the relative rarity of the described random phenomena. This is related both to events with "known" theoretical probability, such as getting a royal flush, and to perceived random events, which are rare based on statistical occurrences, such as earthquakes and being struck by lightening. Theoretically, getting a particular card from a deck of cards has the same features of randomness as getting "heads" when flipping a coin. However, examples like the latter were not present in this pile.

A feature related to relative rarity of the event was that of accident or disaster: the examples mentioned earthquakes, tsunami, being struck by lightening, airplane crush, or dying of a rare disease. Only three examples referred to "lucky" events: a desirable hand in a card game (2) and finding money on the ground. No one mentioned winning a lottery. Further, pushing rarity to the extreme, three examples described events that under normal conditions would be considered impossible:

- Being able to win a marathon with no training
- Walking on a crack in tiles when your stride length is some constant
- Achieving a "perfect" high-five in a massive crowd of people

The connection between randomness and low probability, either calculated or perceived, should be investigated further as it was not featured in prior research.

Study 2: Discussing Randomness with Interviewer In the data presented above, there was no opportunity to ask students for further clarifications—for example, we were not able to ask them why their given examples were random. As pointed out above, many of their examples are subject to multiple interpretations; for example, what notion of randomness does this response entail "forest fires caused by lightning"? Is it the time that is considered random or the location of the next forest that will be on fire by lightning? Would it be non-random if the forest fire was caused by other reasons?

In order to probe in more depth learners' discourse of randomness, we conducted several clinical interviews with undergraduate students, three of which are analyzed in this section. In the interviews, the participants were initially asked to define randomness and give examples and non-examples of random phenomena. Then, they were asked to explain why they considered those examples random. They were also asked to comment on the differences, if any, between everyday randomness and mathematical randomness. At the end of the interview, they were asked whether they could give examples of situations that are more or less random than their previous examples. The interviews were video recorded so that both speech and gestures were captured and then transcribed.

In the analysis of the interviews, we focus on the participants' different ways of communicating randomness. We highlight features of the randomness that have arisen both in the literature review and in Study 1. We attend closely to the participants' non-verbal ways of communicating about randomness, with particular emphasis on their use of gestures.

Kevin: What's Predictable for You May Not Be Predictable for Me Kevin is a prospective teacher with background in biology. When prompted for examples of random events, Kevin responded offering the following two examples: "sneezing for persons without allergies" and "blinking for some people perhaps." In the following exchange, Kevin associates things that have a higher probability of happening as being less random:

I5: OK, so for people with higher probability of sneezing, would then sneezing be more random for these people or less random?
K6: I would say less random.
I7: So you are saying that the higher the probability, the less random it is?
K8: Yes, I would say so.

Consequently, low probability is associated with the *relative rarity* of the event. Further, he offers the seeing of meteors as another example of random events, this time alluding to his inability to predict such an event:

K10: I guess may be from my perspective, of seeing meteors. If someone that has the science calculates how much stuff is in the air and how much of the sky you can see and probably can figure out the probability, but to me it would be pretty random.

When asked directly whether "seeing a meteor is random because it is rare and it happens only a few times in one's lifetime," Kevin emphasizes less the predictability of the meteor passing in the sky than the *unpredictability* of the geographic and temporal location of the seer.

K12: No it is the timing, because you can't really predict when are you seeing it. There is a lot of meteors out there and they do cross the hemisphere just you need to be there at the right place at the right time. You can't predict when are you going to see it.

The interview next asked Kevin to "give an examples of an ordinary English language use of the word random." Kevin explicitly associates random with *unpredictability*, but also evokes its *unexpectedness*.

K14: I think the way it is used in common English, the way I use it is unpredictable[1] (Fig. 1, gesture). For example, when I see someone that I don't expect to see I would say

[1] We will underline words that are accompanied by gestures and describe in parentheses the gesture made.

Fig. 1 The open palms gesture

Fig. 2 Left arm moving to the far left of body

that was random. Maybe if I had thought about it, I would figure out something in terms of probability or something, but in everyday usage if you are not thinking about how often you see something and you see it, makes it all random.

In his description, Kevin also associated the everyday use of the word random with that of *ignorance*, which can be seen both in terms of "not thinking" but also in the gesture used when he says "unpredictable." When describing the meaning of "random" in the phrase 'random guy,' Kevin evokes the *typicalness* sense of random, referring several times to "average person":

K16: I guess when I use the term random guy it is to describe someone that I don't know. An average person (Fig. 1, gesture), someone that is not overly tall or really short, and nothing that sticks out. That would be average (Fig. 1, gesture) and someone that me and the person I'm talking to don't know. So if I'm talking to my friend and I say a random guy, I could say an average person (Fig. 1, gesture) instead.

A slightly different meaning emerges when Kevin describes "random thoughts," which he sees as deviating from the usual:

K20: I think in that situation random would be the thought that wouldn't follow the preceding. If I am thinking about eating, eating, eating and something comes to my mind completely different (Fig. 2, gesture), probably it would be random in that situation.

His gesture evokes the idea that the random though is not in line with the preceding thought, being out of the ordinary. Finally, when asked about whether he thinks of that "absolutely everything is random" or that "nothing is random," Kevin evokes a sense of randomness related to *lack of structure*.

K22: I lean more toward deterministic, if you could crunch all numbers, then you would probably figure out the probability of something happen, I mean given that all of the variables and stuff are known.

Fig. 3 Left hand moving from up to down enclosing/drawing a closed box type space in front of her

Though he never talks specifically about mathematical probability, Kevin's sense of the specific uses of random in everyday language refer to unexpectedness (seeing someone), typicalness (random guy) and relative rarity (random thought). More generally, he speaks of randomness primarily in terms of unpredictability, but also in terms of ignorance and lack of structure. The more things are predictable, the less they are random. Moreover, predictability is a subjective notion (things will be random to him if he can't predict the outcome, even if others can).

Samantha: The More the Number of Possible Outcomes, the More It's Random

Samantha is an undergraduate student majoring in Health Sciences. Samantha was initially reluctant to provide an example of random:

> S2: Something unexpected, out of ordinary. Right? Random in what context? Like any?

When told that she could describe any context, she asks whether we are talking about the "mathematical context" (gesturing with both hand to the left side of her body) or "randomness in life situations" (gesturing with both hands to the right side of her body). The interviewer asks for both. Samantha responds as follows:

> S6: Um, I think, um, in a way they are too similar as to the meaning of random, like what I said random means unexpected (Fig. 3, gesture) and the outcome you can't predict (moves hands to left like before when talking about the mathematical context) whether it is mathematics or whether it is life situation.

The word "predict" seems to be associated for her with the mathematical context, given the gesture. She describes random in terms of *unexpectedness* and *unpredictability*.

When prompted again for examples in both contexts, Samantha pauses for a long time and then asks for help. The interviewer invites her to "think of a sentence in which you would use the word random." When again no answer is given, the interviewer proposes she describes what is meant by "random guy" and "random thought."

> S11: Oh, I see, I would use the word random like I am walking in a mall and somebody approaches me and says something completely out of ordinary, just he doesn't know me, I don't know him, and he asks me something that I don't know the answer to it and I never expected him to ask me, that would be completely random.

Here Samantha alludes to the *lack of structure* aspect of random, particular in terms of the expectations. What sticks out in this example is that the lack of relevance

Fig. 4 Both hands drawing moving arcs

Fig. 5 Fingers open hand moves away from body

Fig. 6 Hands enclose a container

between different pieces of the set of objects/actions described is what makes it random. She also alludes to the *lack of structure* in her description of random thoughts:

> S13: I have those all the times actually, I think about random things (palms parallel, moving right, left, right) when I'm on the skytrain, like how my thoughts just travel (Fig. 4, gesture) from one thing to another thing but I can't, or I don't plan these thoughts. That happens all the times (Fig. 5, gesture) when I'm on a train. Because I have one whole hour (Fig. 5, gesture) to think about anything (Fig. 5, gesture) and I don't plan for what I'm going to think about (Fig. 5, gesture), for example I think about trees (Fig. 5, gesture) and then from trees I randomly jump on my textbook (Fig. 4, gesture) and from my textbook randomly moving to what am I going to eat for dinner. Does that work?

The interviewer prompts Samantha for a mathematical example of random, to which Samantha responds that she took only one mathematics course at the university and in that course "we didn't talk about randomness, we talked about probability, but we never talked about if it means random." The interviewer asks whether she could consider the outcome of the roll of a die as random. When she concurs, the interviewer asks what makes it random.

> S19: It's unpredictable, you have a set of possibilities (Fig. 6, gesture) in this case, six different possibilities, but each of those possibilities are unexpected and unpredictable (hands still holding the container making beat gestures at the borders of the container), so each role would be random, I guess. OK, for example, say playing roulette, that would be a game that is random because every time the person spins the ball it is random, the number that the ball will land on is random (Fig. 5, gesture).

Samantha speaks again of *unpredictability* and *unexpectedness*. But she also emphasizes the way in which randomness is related to having a fixed range of possibilities, which she expresses as well through her container gesture.

Fig. 7 Fingers tight hand moving from up to down, as if putting a stop to a moving thing

Fig. 8 Palm up hand moving slightly away from body

The interview then prompts Samantha for "examples of life that are non-random."

S21: Non-random would probably be like let's say I'm going to school, all of the events that occur as I'm going to school are not random (Fig. 7, gesture). Because they are planned (Fig. 7, gesture). I would get on the skytrain (Fig. 7, gesture) and then I would get on the bus (Fig. 7, gesture) and I would get into my class (Fig. 7, gesture), like all these events are planned (Fig. 7, gesture) they are predictable (Fig. 7, gesture with both hands).

Here Samantha contrasts randomness with things that are planned and predictable. Her repeated gestures evoke also structure and control, which resonates with the *unsteerability* notion of the randomness. The interviewer then asked Samantha to give examples of "events with 100 % probability of happening."

S25: I think, it happens with my drawer of socks, because I have a drawer of socks that are all white because I am very organized so I put all of my white socks in one drawer, and all of my colored socks are in another drawer (gesturing a container first to the right and then to the left) so 100 % of the time that I go to my white sock drawer, 100 % of the time I will pull out a pair of white socks.
I26: Do you see any randomness in this experiment of socks drawer?
S27: There could be, there could be the possibility that my mother put the wrong sock in the wrong drawer (Fig. 8, gesture) and then I might have a chance of pulling black socks from the white sock drawer (Fig. 8, gesture).

The interviewer expresses some surprise at the idea that Samantha just puts her hand in and grabs what comes out, to which Samantha responds as follows, using the idea of random as rare, contrasting it with *typicalness*.

S29: (laughs), yes. Well, I mean when I look at my white sock drawer it is random to see a black sock, it's not supposed to be there. If there is, that would be random.

At this point, the interview asked Samantha about whether there are things that are more or less random. After a long pause, the interview mentioned the rolling a die situation and asked whether there was "another experiment that is more or less random than this one?"

S32: I think if there was more dice, more than one, um, actually it doesn't work, umm, let's say if there is five dice instead of one, the outcome would be different (Fig. 8, gesture)?

Fig. 9 Both hands enclosing a long container

> Not sure. It would be more unpredictable (Fig. 8, gesture)? The outcomes are larger (Fig. 9, gesture) versus one die when we have six outcomes than if we put more dice (Fig. 9, gesture) it would be 12, 36, and so on, so that would be more random vs. less random, maybe?

In contrast to Kevin, who associated the amount of randomness with the predictability, Samantha associates the range of possible outcomes with randomness. Her gestures indicate the size of the sample space; the bigger the size, the more the situation is random—this kind of thinking is similar to the 'the more X, the more Y' intuition, described by Stavy and Tirosh (2000). Moreover, a large sample space is also connected to a *relative rarity* of each particular event, assuming even distribution.

In order to further inspect Samantha's understanding of random, the interviewer asks her to comment on whether HTHTHT is a less probable event compared to other outcomes since it has a pattern in it and it is not very likely that a coin produces a pattern.

> S34: I think that I agree, because when you flip a coin it is 50–50 each time either this or that, it is not like the coin will land straight on the edge, so I don't think that there is too much randomness in flipping a coin.

When asked whether she thinks that a die has more randomness in it since it has six sides, Samantha responds "Yes, more randomness if we increase the number of dice." Although Samantha seems unsure about describing and exemplifying randomness, she certainly associates it with being about unpredictability and lack of structure and, in comparing it with non-random, with unsteerability. But she also contrasts it with typicalness. Further, situations are more random the more possible outcomes they have, which is connected to relative rarity. Samantha talks about both random and non-random situations in terms of a range of outcomes; but random events are different because the outcomes can be different.

Tyler: Maybe You Can Predict It but You Can't Accurately Predict It Tyler is an undergraduate student majoring in mechanical engineering. He is first asked for examples of random things.

> T2: Sure, I suppose everything has a degree of randomness to it. What seat you choose on a bus, what time you head for, you know, school in the morning, how much milk I pour (gesturing the pouring of milk) in this coffee, it has elements of randomness to it and, of course, the obvious ones: playing dice or cards, or you know (Fig. 10, gesture) any sort of video games, RPGs specifically when you are attacking (right hand makes small quick pushes back and forth in front of his face) these little random things (points with whole hand to what looks likes an ascending set of stairs with the hand is holding a small object). Even when I'm sipping a coffee the amount of coffee, I'm sipping is random because it is not the same every time.

Tyler's sense of randomness relates here, especially when he talks about the sipping of coffee, to *variability*. When the interviewer asks Tyler to explain why "not being

Fig. 10 As if throwing something to the back

Fig. 11 Gesture space moves to his right side of the body as he speaks about real world

the same every time" is an example of random, Tyler explains that it comes down to a matter of "belief" since there are "two views on how this world works" (the deterministic and the non-deterministic). In the former world, he explains:

> T4: [I]f you know all the information about the system you could accurately predict it; meaning it wouldn't be random at all, however (Fig. 11, gesture), in our world it doesn't really work that way because when you get really, really close down like to quantum level to the atoms [...], it is impossible to know the position of atom as well as its direction and velocity and momentum [...] at the same time [...], it is going to have some level of randomness to it.

Tyler continues on to explain how the way a drop of water will fall on your hand "is random because depending on how deep it sits on your hand on the molecular level, it might go the other way." At the quantum level, he describes the behavior of the "is unpredictable," and asks "and so how can the system as a whole be completely predictable?" Later, he continues on to describe the way in which "when I'm taking a sip of my coffee there is randomness (hand palm open loosely falls down to his lap) to it and it is almost impossible to say how much exactly."

Up until now, Tyler's sense of randomness is related to *unpredictability*, but his talk about the quantum level also involves a level of sense *ignorance* since operations there are unknowable. The interviewer then asked for other examples of randomness and he talks about the exact layout of the table in the room and where a garbage can is placed, saying that "Everything I believe has a degree of randomness to it even if the tiny, tiny degree." He then returns to the idea of *variability*.

> Consider M&M's in a bag. M&M's are supposed to be made of the same shape (Fig. 12, gesture), but if you look at each of them, you see that one has a little lump to it, other has a distortion to it and, however, they are manufactured the same way but still there is some

Fig. 12 Fingers enclosing a round shape

Fig. 13 Both hands enclosing a container

Fig. 14 Moving right hand toward the left and covering the previously indicated set of things

degrees of variation (left hand moving from elbow up to down, with wrist rotating as if drawing) between each one.

The interview prompts Tyler to talk about the use of the word random in common language, mentioning in particular "random thoughts" and "random guy." For the latter, he describes a random guy as being "out of the normal" and "going against the crowd." He elaborates that the use of the word 'random' in natural speech is:

> T10: [...] a little different than its actual meaning, they just mean different, weird, and out of ordinary whereas random by definition means unpredictable or means this many (Fig. 13, gesture) and it is going to be one of them (Fig. 14, gesture), but we don't know which until you choose. And again about the random thoughts I think the way they use it in speech it doesn't pertain to what they are doing at the moment. If you are playing soccer and you think about your mathematics homework, it is random because it doesn't deal with what you are doing at the time.

In describing unpredictable, Tyler refers with speech and gesture to the range of possibilities an event can have and to the randomness being linked to the fact that the outcome is not known (until it is chosen). But both examples of non-mathematical uses of the word 'random' relate to *lack of structure*. Tyler also joins Samantha in describing the everyday use of randomness in terms of *not* being typical. This is different from Kevin who associated randomness with typicality.

The interviewer further prompts Tyler on his definition of probability in terms of unpredictability. Tyler elaborates in more mathematical terms, referring to random events, probability, and certainty:

> T12: I guess, you can predict random events (Fig. 13, gesture but with palms face down) but you can't say which one for certain it is before it happens. If something is random, you don't know exactly what is going to happen, maybe you can predict it, but you can't

accurately predict it, like with 100 % certainty. We can say it is going to be probably one of these ten things (repeating above gesture), randomly one of them (Fig. 14, gesture). So, it could equally be either one (vertical hand movements that partitions the set), or I suppose it doesn't necessarily mean equally but depending on wording I suppose it is implied sometimes?
I13: I don't know, let me ask you this, you said that randomness means not 100 % certain, can we come up with examples of randomness with very, very high probability, not 100 % though?
T14: Sure, like the probability of flipping heads with a coin for 100 times in a row, it is possible; but it is very, very random because it doesn't happen very often.

The interviewer asks whether this probability is low and Tyler response "Yes, what I mean is the probability of this not happening. 99.9 % is probability of it not happening." He continues on to say the event is "very rare but it still might happen, it is random and if it did happen I would say it is very random." This is the most evident example of connecting randomness to the *relative rarity* of an event.

On Gestures that Accompany Randomness The sources of randomness identified by the interviewees are similar to, though somewhat more diverse than, those of the examplification data. For example, both emphasize *unpredictability* as a defining feature of randomness. The gestures used by the participants further underscore the way in which the *unpredictability* features strongly in their discourse of randomness. The hands wide open, palm up gesture co-occurs with the word random and seems to communicate a lack of control over the situation, or a lack of knowledge.

Another noticeable aspect of the interviewees' discourse on randomness relates to their strong distinction (expressed through gestures—hands moving to one side, and then to the other; pointing to one side or the other) between everyday randomness and mathematical randomness. Prior research has not drawn attention to the ways in which learners are aware of differences between the way the word 'random' is used in each context or to the distinctions learners attribute to each type of usage. Further, by specifically exploring everyday uses of the word random, we were able to see how the everyday discourse is used in trying to participate in the mathematical discourse so that ideas such as typicality and relative rarity, which are relevant to everyday use, also come to be relevant to mathematical use.

Finally, two of the interviewees made a container-like gesture in several instances when referring to the possible events or outcomes of a phenomenon. Although not explicit in their speech, the gesture suggests that the participants think of the possible outcomes as belonging to a closed, finite set. This seems particular relevant to the discourse of mathematical randomness in the sense that one needs information about the distribution of the possible outcomes in order to say something about whether a particular sequence is random. We suggest that this gesture might be helpful to evoke in a didactical situation.

5 Concluding Remarks

The notion of randomness is central to the study of probability and statistics. However, it is rarely explicitly explained, despite the continued efforts of mathematicians to produce a rigorous definition. Samantha's comment seems to illustrate the issue: "We didn't talk about randomness, we talked about probability, but we never talked about if it means random." Researchers in mathematics education have identified some intuitively-constructed features of randomness. However, most studies relied on students considering sequences of outcomes of events, focusing on binary outcomes, such as a coin toss. Our research broadens these approaches, moving away from the context constructed by researchers towards examples and descriptions provided by participants. We focused on (i) how adult learners (prospective secondary school teachers and undergraduate students) exemplify random phenomena and what features of randomness are present in their examples, and (ii) how learners communicate ideas of randomness through a broad multimodal discourse.

We first attempted to categorize participants' written examples of randomness using the three types suggested by Tsonis' (2008) model. We found that such classification significantly depends of participants personal knowledge and experience, rather that on the described event. We then attempted to classify participants' examples using features of randomness identified by Pratt and Noss (2002). We noted that many examples illustrate more than one feature, as there is a significant overlap among unpredictability, unsteerability, and irregularity. While these features help characterize participants' intuitions, they do not provide sufficient information on how the notion of randomness is *used* by learners. As such, we extended our study through the use of clinical interviewers in which participants' phrasing and gestures could be analyzed.

The following attributes of randomness were featured in participants' responses, extending and refining the list previously compiled by Pratt and Noss (2002): unpredictability, unsteerability, unexpectedness, lack of structure, variability, ignorance, typicalness, and relative rarity. The most common gesture that accompanied the discussion of randomness was that of open hands with palms up, which often communicates in daily conversations the "I don't know" phrase. It was strongly related to the features of unpredictability and ignorance in participants' talk, which were prominent in this group of participants. Unexpectedness, lack of structure, and variability are related to the feature of irregularity, previously identified by Pratt and Noss (2002).

However, the notion of relative rarity was not mentioned in prior research, but appeared repeatedly both in the interviewers and in the examples provided in writing. That is, events that have low probability of occurring were considered by our participants as random. This was the case both for mathematical randomness and for randomness in everyday usage of the word. Further, the idea of relative rarity, which can be seen as contradicting the idea of typicality, actually appears to co-exist with it. It all depends on the attributes that are associated with randomness.

Based on our findings, we suggest that teachers (and teaching materials) attend more explicitly to learners' discourse on randomness and, in particular, to learners' everyday uses of the word. These can be compared and contrasted with different definitions of randomness available in the mathematics discourse. It seems that the notion of randomness that is related to the passing of certain statistical tests might be the most distinct from learners' discourse on randomness—this very pragmatic view of randomness seems particularly amenable to learning situations involving computer-based technologies.

References

Arzarello, F., Paopa, D., Robutti, O., & Sabena, C. (2009). Gestures as semiotic resources in the mathematics classroom. *Educational Studies in Mathematics*, *70*(2), 97–109.

Bailey, D., Borwein, P., & Plouffe, S. (1997a). On the rapid computation of various polylogarithmic constants. *Mathematics of Computation*, *66*(218), 903–913.

Bailey, D., Borwein, P., & Plouffe, S. (1997b). The quest for Pi. *The Mathematical Intelligencer*, *19*(1), 50–57.

Batanero, C., & Serrano, A. (1999). The meaning of randomness for secondary school students. *Journal for Research in Mathematics Education*, *30*(5), 558–567.

Bennett, D. J. (1998). *Randomness*. Boston: Harvard University Press.

Chernoff, E. J., & Zazkis, R. (2010). A problem with the problem of points. In P. Brosnan, D. Erchick, & L. Flevares (Eds.), *Proceedings of the thirty-second annual meeting of the North-American chapter of the international group for the psychology of mathematics education* (Vol. VI, pp. 969–977). Columbus: Ohio State University.

diSessa, A. (1993). Towards an epistemology of physics. *Cognition and Instruction*, *10*, 105–225.

Durand, B., Kanovei, V., Uspenski, V., & Vereshchagin, N. (2003). Do stronger definitions of randomness exist? *Theoretical Computer Science*, *290*, 1987–1996.

Falk, R., & Konold, C. (1998). Subjective randomness. In S. Kotz, C.B. Read, & D. L. Banks (Eds.), *Encyclopedia of statistical sciences* (pp. 653–659). New York: Wiley.

Fischbein, E., & Schnarch, D. (1997). The evolution with the age of probabilistic, intuitively based misconception. *Journal for Research in Mathematics Education*, *28*(1), 96–105.

Kahneman, D., & Tversky, A. (1972). Subjective probability: a judgment of representativeness. *Cognitive Psychology*, *3*, 430–454.

Li, M., & Vitany, P. (2008). *An introduction to Kolmogorov complexity and its applications*. New York: Springer.

Marsaglia, G. (2005). On the randomness of Pi and other decimal expansions. *Interstat*, *5*.

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

Sinclair, N., & Gol Tabaghi, S. (2010). Drawing space: mathematicians' kinetic conceptions of eigenvectors. *Educational Studies in Mathematics*, *74*(3), 223–240.

Stavy, R., & Tirosh, D. (2000). *How students (mis)understand science and mathematics*. New York: Teachers College Press.

Strzalko, J., Grabski, J., Perlikowski, P., Stefanski, A., & Kapitaniak, T. (2009). *Lecture notes in physics: Vol. 792*. *Dynamics of gambling: origins of randomness in mechanical systems*. Berlin: Springer.

Tao, T. (2008). *Structure and Randomness*. Providence: American Mathematical Society.

Tsonis, A. (2008). *Randomnicity, rules and randomness in the realm of the infinite*. London: Imperial College Press.

Vitanyi, P. (1994). Randomness. In A. Schrijver, N. Temme, & K.R. Apt (Eds.), *From universal morphisms to megabytes: a Bayesian space odyssey* (pp. 627–642). Amsterdam: Math. Centrum, Centrum Wisk. Inform Press.

Watson, A., & Mason, J. (2005). *Mathematics as a constructive activity*. Mahwah: Erlbaum.

Wolfram, S. (2002). A new kind of science. Retrieved February 10, 2012, from Wolframscience website: http://www.wolframscience.com/nksonline/toc.html. Wolfram Math world available at: http://mathworld.wolfram.com/RandomNumber.html.

Zazkis, R., & Leikin, R. (2008). Exemplifying definitions: a case of a square. *Educational Studies in Mathematics*, *69*(2), 131–148.

Developing a Modelling Approach to Probability Using Computer-Based Simulations

Theodosia Prodromou

Abstract The introduction of digital technology into secondary schools is ideally suited for supporting students as they manipulate and portray data in a range of different representations to draw inferences from it without relying on a classical understanding of probability theory. As a result, probability is overlooked from school curricula and is gradually becoming almost a non-existent topic. The aim of recent curricula (e.g. ACARA 2010) to support the parallel development of statistics and probability and then progressively build the links between them seems utopic since statistics prevails over probability in mathematics curricula. In this chapter, it is argued that it is worthwhile to consider an alternative approach for teaching probability—presenting probability as a modelling tool, which reflects the mindset of an expert when using probability to model random behaviour in real-world contexts.

This chapter discusses results from two recent research studies (Prodromou 2008, 2012) that investigated middle school students' understanding of probability as a modelling tool. Within this chapter, the students' reasoning will be considered from different perspectives:

(1) How students articulated fundamental probabilistic concepts associated with the construction of univariate probability models when using probability to model random behaviour.
(2) Students' discussion as they engaged in exploring recently developed computer-based simulations which treat probability as a modelling tool.

In this chapter, references will be limited to students in Grades 6 to 9. Prodromou's research studies address four research questions as follows: (1) How do middle school students use probability to model random behaviour in real-world contexts? (2) What connections do they build among fundamental probabilistic concepts when treating probability as a modelling tool? (3) How do they synthesize the modelling approach to probability with the use of distributions while concurrently

T. Prodromou (✉)
University of New England, Armidale, New South Wales, 2351, Australia
e-mail: tprodrom@une.edu.au

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_22,

making inferences about data? (4) What activities can be designed to support the proposed alternative approach for teaching probability?

The results of this study provide answers to the aforementioned research questions and suggest that the way students express the relationship between signal and noise is of importance while building models from the observation of a real situation. This relationship seems to have a particular importance in students' abilities to build comprehensive models that link observed data with modelling distributions.

1 Introduction

At a secondary level, there are two main approaches to teaching probability: classical (or classicist) probability and relative frequency.

Before 1970, the classical view of probability, which is based on combinatorial calculus, dominated the secondary school curriculum. The Classicist approach had emerged in the early eighteenth century; in it probability was defined by De Moivre (1967/1756) as a fraction of which the "Numerator is the number of chances whereby an event may happen and the denominator is the number of all chances whereby it may either happen or fail" (De Moivre 1967/1756, p. 1). Pierre-Simon Laplace (1995/1814), in his *Essai Philosophique sur les Probabilités* published in 1814, made an implicit assumption of equal likelihood of all single outcomes of the sample space, and thereby defined probability as the fraction of outcomes favourable to this event in the sample space. This classicist perspective introduced an a priori approach to probability that means that probabilities can be calculated before any physical experiment is performed.

At the end of the twentieth century, the introduction of digital technology into schools prompted interest in Exploratory Data Analysis (EDA; Tukey 1977), thus there was an evolution of mathematics teaching towards more experimental activities. This led to another fundamental epistemological choice for the introduction of the notion of probability. This epistemological choice is based on observations of relative frequencies of an event associated with a random experiment that is repeated a sufficiently large number of times under the same conditions. In this view, probability is estimated as a limit towards which the relative frequencies tend when stabilizing (von Mises 1952/1928). The idea of stabilization is based on the empirical laws of large numbers. This is called the "frequentist" approach to probability (Renyi 1992/1966). Thus the notion of probability can be introduced through these two different approaches, either by means of relative frequencies or by means of the classical definition.

The modelling approach to teaching achieves a co-ordination between these two perspectives. Chaput et al. (2011) refer to the modelling perspective not as a distinct third approach that contrasts with the classicist and frequentist approaches to probability, but as an attempt to synthesize them. Recent research (Prodromou 2008) showed that this synthesis is better achieved by the use of simulations. Recent curricula studies (Chaput et al. 2011) have adopted this modelling process using simulations of models in stochastics (probability and statistics).

Such an approach requires further research in students' understandings.

2 Probability in the Curriculum

In current curricula, it seems that the choice is often made to present probability as either theoretical (the classicist approach) or experimental (the frequentist approach). But some curricula choose to blend them. For example, the Mathematics K-10 Syllabus in Australia strand has focused on the collection, organization, display and analysis of data, as well as the calculation of theoretical and experimental probabilities of simple and compound events.

In particular, by the end of the primary school, students are expected to (i) recognise that probabilities range from 0 to 1; (ii) describe probabilities using fractions, decimals, and percentages; (iii) conduct chance experiments with both small and large numbers of trials using appropriate digital technologies, and (iv) compare observed frequencies across experiments with expected frequencies. At the primary level, the curriculum in Australia encourages students to make judgements related to subjective probabilities, when attempting to associate 0 with the impossible, 1 with certainty, and values between 0–1 with varying degrees of uncertainty. Subjective probabilities are left behind (without any reference to them) once students are introduced to the classical approach to probability (ii). Students are expected to conduct chance experiments in the classroom using coins, dice, spinners, or any other kind of random generator to perform small numbers of trials (iii). Students are also expected to use appropriate digital technologies to show how the experimental probability approximates the theoretical probability when the experiment is performed a large number of times under the same conditions (iv).

At years 7–9, the national curriculum has focused on (i) constructing sample spaces for single-step experiments with equally likely outcomes; (ii) assigning probabilities to the outcomes of events and determine probabilities for events; (iii) identifying complementary events and using the sum of probabilities to solve problems; (iv) describing events using the language of 'at least', exclusive 'or' (A or B but not both), inclusive 'or' (A or B or both) and 'and'; (v) representing events in two-way tables and Venn diagrams and solving related problems; (vi) listing all outcomes for two-step chance experiments, both with and without replacement using tree diagrams or arrays; (vii) assigning probabilities to outcomes and determining probabilities for events; and (viii) calculating relative frequencies from given or collected data to estimate probabilities of events involving 'and' or 'or'. At the secondary level, the National Curriculum in the Australia expects students learn how to build an event space via combinatorial analysis and to use algebra of probabilities and representations ((vi) and (vii)) to achieve (viii).

Although the aim of the national Australian curriculum (ACARA 2010) is to support the parallel development of statistics and probability and then progressively build the links between them, at present statistics are given the bulk of attention for several reasons. On the one hand, the multiple applications of probability to different sciences were hidden and probability was considered as a subsidiary part of mathematics, since it is too tightly connected to games of chance and only dealt with such games. This tight connection of probability with games of chance was possibly perceived as amoral. On the other hand, the increasing interest in statistics and the growing expectations around what it is expected from students at the school level when handling data are changing rapidly. The new techniques in Exploratory Data Analysis (Tukey 1977) and continuing developments in pioneering software used in statistics education, such as Fathom (www.keypress.com/x5656.xml) and *Tinkerplots* (www.keypress.com/x5715.xml), are ideally suited for supporting students as they manipulate data and portray it in a range of different representations to draw inferences from it reducing the need for sophisticated understanding of probability theory.

EDA is developing interesting pedagogic approaches towards informal inference (Tukey 1977). EDA was developed as a means of engaging students in data analysis; it promotes an approach to interpreting data using multiple graphical representations and summary statistics to see the underlying structure of the data. In terms of the classical approach, the sample is not assumed to be random or a sample from a larger population. In essence, EDA is concerned with whole populations or unique data sets. In particular, a focus on comparing data and focusing on the group propensities of data have been seen as particularly productive for supporting students in making sense of data in a meaningful way.

The novel technological developments together with EDA techniques have prompted researchers' interest in students' informal inferential reasoning (IIR), a learning process of drawing conclusions from data and (i) generalising beyond data available (e.g. parameter estimates, conclusions, and predictions); (ii) using data as evidence of the generalisation; and (iii) articulating the degree of certainty (due to variability) embedded in a generalisation (Makar and Rubin 2009, p. 15). Introducing statistical concepts at an informal level to students from the early school years carries with it a potential to change the focus of statistics from the calculations and graphing techniques that usually define statistics in schools (Sorto 2006) to a revolutionary way of reasoning behind statistics (Makar et al. 2011). This change in the focus of statistics supports students in building stronger connections between statistics and data from real world phenomena through long-term everyday experiences. By strengthening links between statistics and everyday experiences through repeated and long-term exposure to inferential reasoning, children will acquire deep conceptions of data before formal ideas will be introduced. Making informal inferences does not require any recourse to a classical understanding of probability. The third characteristic of the informal statistical inference, however, emphasizes that inferences encompass uncertainty and their degree of uncertainty should be fully expressed. For example, when estimating a population mean from a sample, this estimate must be made acknowledging its inherent uncertainty. The expression of such an uncertainty, however, does not lead to any probabilistic articulations.

While the statistics curriculum has adopted new approaches like EDA, and new technologies such as Fathom and *Tinkerplots*, and is changing in an attempt to adapt to current software innovations and the enquiry-based stance for learning and teaching statistics, the probability curriculum remains isolated in its world of dice, spinners, and coins, which contributes to its neglect in many classrooms. The modelling approach provides a response to the reasons that probability has been vanishing from some international mathematics curricula (Borovcnik 2011).

Recently, some international curricula, for instance, in France, have opted for the modelling approach (Chaput et al. 2011) when teaching probability. This approach reinforces use of models that are formalised in a symbolic system and developed to represent concrete situations or problems arising from reality. Such modelling of concrete situations is a process of building models that cannot always be deterministic. The models that incorporate uncertainty or random error in a formalised way are called probabilistic models. These probabilistic models, according to their inherent rules, are expected to simulate the exact behaviour of random phenomena and also predict specific properties of random phenomena. For instance, random generators can be seen as determined if only one was aware of a set of factors that causally affect the behaviour in which case the outcomes would be entirely predictable. In practice, this is an unlikely state of affairs and it is likely that one would be interested instead in adopting a probabilistic model.

In this framework, a probability distribution of some discernible characteristics has the status of a model of the data that describes what one could expect to see if many samples were collected from a population, enabling us to compare data from a real observation of this population with a theoretical distribution. This perspective is in accord with the *modelling* process of *contemporary* statistical thinking (Wild and Pfannkuch 1999) that allows the application of theoretical results when making statistical inferences regarding particular observed sample statistics or data analysis.

The modelling approach points out that probability should not only be used to explain the methods of inferential statistics at the high school level but also "as a tool for modelling computer-based action and for simulating real-world events and phenomena" (Pratt 2011, p. 3). Pratt argued that "it is necessary therefore to stress in curricula an alternative meaning for probability, one that is closer perhaps to how probability is used by statisticians in problem solving" (2011, p. 1). Pratt offers examples of recent developments in software that provide the user with modelling tools based on probability (i.e. the user can set the probability of some event) that could be used to construct models used by computer-based simulations.

One of the examples that Pratt gives is the work of Prodromou (2008), who investigated how 15-year-old students were able to make connections between a data-centric perspective on distribution and a modelling perspective on distribution. Based on the secondary curriculum with its focus on using distributions to make inferences, it was anticipated that students might articulate a data-centric perspective on distribution that would be consistent with the Exploratory Data Analysis (EDA) approach. One alternative view the students might adopt would be to recognise the probabilistic features of distribution. Statisticians often explain variation in data distributions as being either the result of noise or error randomly affecting the main effect, or higher random effects. By contrast, the modelling perspective reflects the mindset of statisticians when proposing models of how the data are generated-out of random error and various effects.

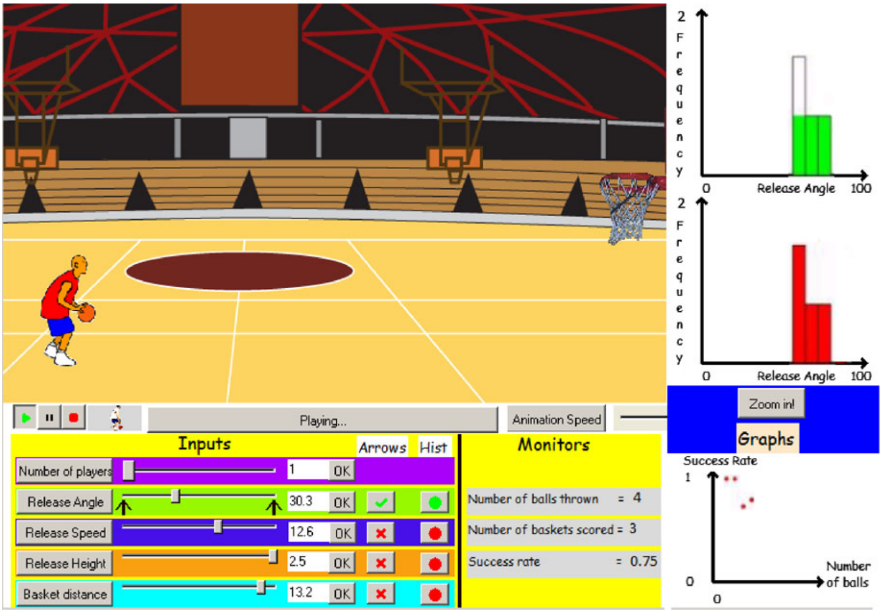


Fig. 1 The sliders allow the student to select how many balls should be thrown, the release angle, release speed, release height and basket distance. The value of the parameter is chosen randomly by the computer from a Normal distribution of values, centred on the position of the handle of the slider and spread between the *two arrows*. Results are displayed in monitors such as number of balls thrown, number of goals scored and success ratio. The graph of success ratio against time by default is presented in the Interface window (*bottom right*). The histogram of the variable that has *arrows* on can be presented in the Interface window (*two top graphs on the right*). The histogram (*middle right*) displays the frequency of release angles used in the basketball throws (*right*) and also how often these were successful throws (*top right*). Similar graphs of any variable can be chosen through an options menu

Prodromou (2007, 2008, 2012) built a computer-based simulation of a basketball player attempting to make a basket. Underlying the simulation were two mechanisms for generating the trajectories of the balls following Newton's Laws of Motion; one was fully deterministic; the other used a probabilistic model that incorporated variation in the trajectories. The interface was designed in such a way that the data-centric perspective was presented graphically as a set of data about the trajectories and success of shots at the basket (Fig. 1), and the modelling perspective was presented as the probability distribution that generated the varying trajectories of the balls. In the data-centric perspective, data are the starting point, spread across a range of values, from which trends or models might be discovered. Prodromou and Pratt (2006) refer to that spread of data in the data-centric distribution as portraying variation. In the modelling perspective, data are the outcomes, generated by random selection from the values of the model. Prodromou and Pratt (2006) refer to spread of values in the modelling distribution as portraying randomness. Prodromou's research was interested in investigating whether and how students co-ordinated the

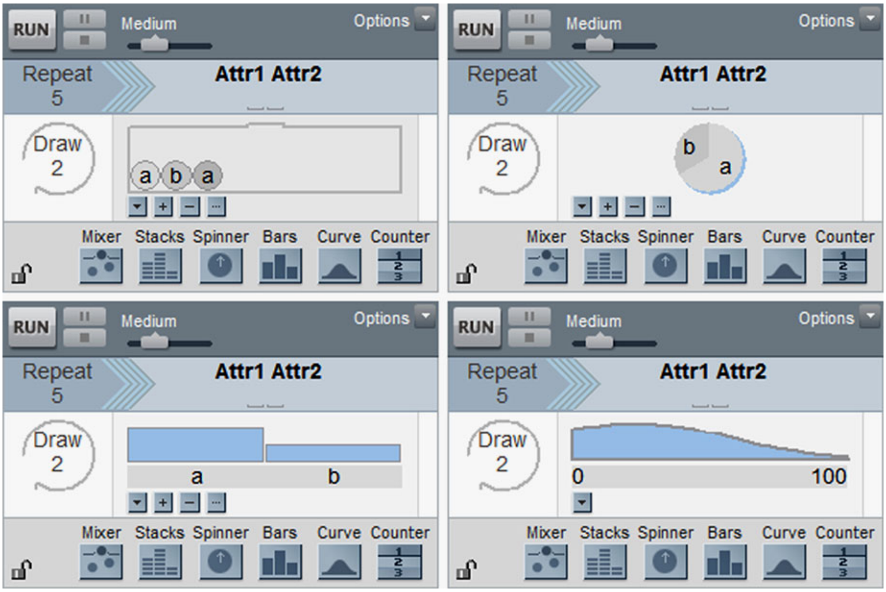


Fig. 2 Examples of samplers in TinkerPlots2

experimental outcomes with the theoretical outcomes produced from a theoretical model.

The second example given by Pratt (2011) is *TinkerPlots*, software that has been developed by a team at the University of Massachusetts at Amherst led by Konold. The *TinkerPlots* software is designed to help students in grades 4–8 develop understanding of data, probability, and statistical concepts. *TinkerPlots* provides students with tools to analyse data by creating a variety of colourful representations that will help reveal to students the structure of real data and recognise any patterns that the data follow as they unfold. The most recent version, *Tinkerplots2*, was used in this study. It offers new tools that exploit probability as a modelling tool using the sampler that is essentially a non-conventional form of probability distribution.

Using *TinkerPlots*2, the student can create a modelling distribution using a mixer (top left in Fig. 2), set up a sampler as a spinner by defining the sizes of the sectors (top right), a histogram by determining the heights of the bars (bottom left), or a probability density function by drawing a curve to define a probability density function (bottom right).

All these options of the sampler engine were used by the grade 9 students in the following investigation of their attempts to use probability as a modelling tool to build a model of a real-world phenomenon because "the most effective way to create process knowledge is to develop a model that describes the behaviour of the process" (Hoerl and Snee 2001, p. 230). This process knowledge, however, will inevitably bring with it new challenges in how children learn and give rise to research questions about the conceptual development of students who will engage in connecting probability to statistics and to simulated real phenomena.

Recent teaching studies have led to teaching based on the modelling approach using simulations of models in connecting statistics to simulated real phenomena, fostering the growth of basic probabilistic concepts. Such a choice will necessarily bring with it new challenges in how children learn and give rise to research questions about the conceptual development of students who will engage in connecting probability to statistics and to simulated real phenomena.

Rather than assessing the efficacy of different tasks that students engaged with in the two studies, this chapter steps back to reflect on how middle school students articulated fundamental probabilistic concepts associated with the construction of univariate probability models when using probability to model random behaviour.

This chapter is part of larger examination of four research questions: (i) How do middle school students use probability to model random behaviour in real-world contexts? (ii) What connections do they build among fundamental probabilistic concepts when treating probability as a modelling tool? (iii) How do they synthesize the modelling approach to probability with the use of distributions while concurrently making basic inferences about data? (iv) What activities can be designed to prompt and support the proposed alternative approach for teaching probability?

The next two sections of this chapter will examine the building of middle school students' understanding when using probability as a modelling tool.

3 First Study: BasketBall

The *BasketBall* simulation promotes a view that learners might adopt by recognising the probabilistic features of distribution. This is consistent with how statisticians explain variation in data distributions as being either the result of noise or error randomly affecting effect, or higher random effects.

We will examine the students' interaction with the fourth iteration of *BasketBall* simulation, in which they were given the opportunity to access to the graphical representation of the modelling distribution, which showed a distribution of values from which the computer would randomly choose. In the *BasketBall* simulation, the modelling distribution mirrors the probability distribution as generating the trajectories of the basketballs. Hence, in the *BasketBall* simulation, the modelling distribution takes on the utility of generating data.

There were two ways students could set the modelling distribution: they could either adjust interface controls (the arrows or the handle on the slider, see Fig. 3), or they could change the way in which values were chosen by directly entering their own value for each outcome interval (Fig. 4). In both cases, the simulation allowed students to transform the modelling distribution (which generates the output data) directly and thus they have indirect control over the data-centric distribution.

The conclusions drawn from the case studies show the construction of three interpretations for bridging the modelling and the data-centric perspectives on distribution:

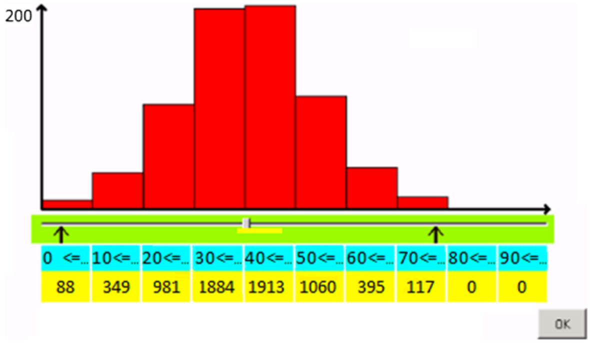


Fig. 3 The simulation affords students the opportunity to alter the modelling perspective, thus change the way that the computer generates the data, by moving either the *arrows* or the handle on the slider

- General Intention (I_G): The modelling distribution (MD) was perceived as the intended outcome and the data distribution (DD) as the actual outcome, suggesting a connection being made, in which the modelling distribution in some sense generates the data.
- Stochastic intention (I_{st}): General intention becomes stochastic intention (I_{st}) when randomness becomes a part of the interpretation of the student.
- Target: The *modelling* distribution (MD) was perceived as the target (T) to which the data distribution (DD) is directed.

When students manipulated the arrows or the handle on the slider of the *Basket-Ball* simulation, some of them appeared to perceive of the modelling distribution as the intended outcome and the data-centric distribution as the actual outcome. Two

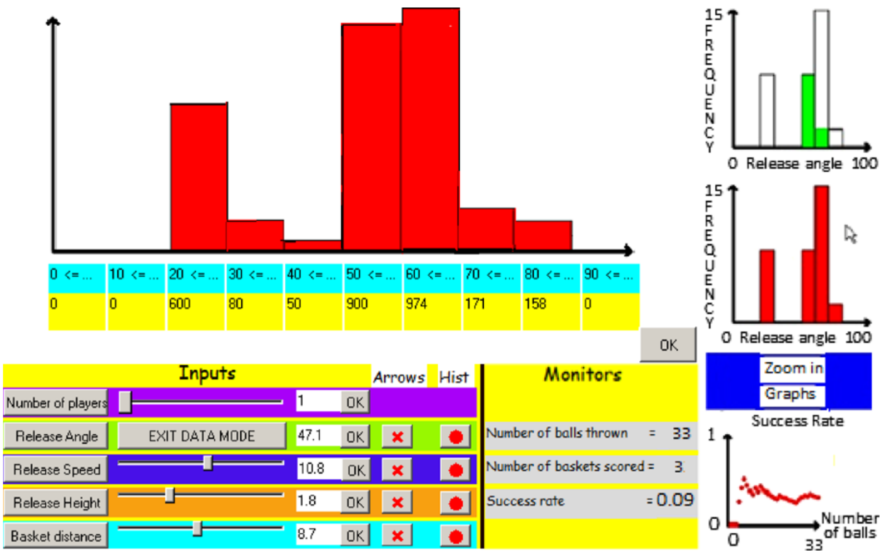


Fig. 4 The simulation affords students the opportunity to alter the modelling distribution directly by setting numerical values associated with each possible outcome for a given variable

pairs of students had a sense of a general intention (I_G) when they talked about variation. Their articulations were explicitly characterised by the absence of a strong sense of the probabilistic mechanism and as evidenced, there was not a progressive articulation of intuitive relationships, leading gradually to a clear probabilistic-type language for talking about randomness and probability. Their reasoning about intentionality remains insufficiently clear. We can have at least two different possible interpretations: (a) the intention is simply an expression of the pre-programmed deterministic nature of the computer—at least in their experience, or (b) intentions are reflected in the actions of a modelling builder.

Three students expressed a sense of a stochastic intention (I_{st}). When they did, they seemed to see the modelling distribution as the intended outcome, progressively generating the data-centric distribution. It can be concluded from their various references to chance that this perception of intention is probabilistic, and underpins the idea that the modelling distribution precedes, indeed generates, the data-centric distribution. Hence, the students refer to the computer 'wanting' to throw the ball at various angles, a sentiment that we characterise as the situated abstraction, "the more the computer wants to throw at a particular angle, the higher is that angle's bar".

Finally, 6 pairs of students made the opposite connection from the data to the modelling distribution by perceiving the latter to be the target to which the former is directed, without understanding the probabilistic mechanism by which the throws were selected randomly from the intervals of the modelling distribution.

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

One pair of students was an exceptional case study. When they manipulated the simulation controls, instead of recognising a semantic similarity between the two perspectives on distribution, they recognised a syntactical similarity (X_{sy}). In fact, they solely paid attention to superficial elements of the two perspectives, such as different sizes of the two graphs. When they entered their own data, the scaling problem of comparing the two graphs proved a significant constraint, especially since students at this age have immature proportionality schema. They consequently struggled to see any similarity between the two distributions. Similarly, two students from two other pairs constructed syntactical connections.

Overall, with the exception of 4 students who managed only syntactical connections, and 2 students who made only a target type connection, all other students (10) were able to articulate at some point both intentional and target type connections.

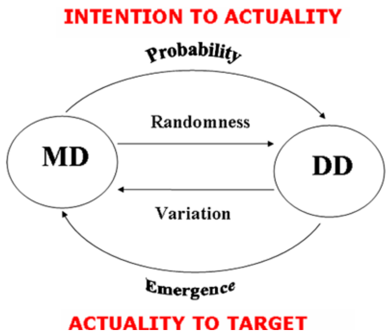


Fig. 5 A tentative model for the connection of the data-centric and modelling perspectives on distribution

The very fact that these students did articulate both types of connection suggests that exercises of the sort described in this study might contribute to successful co-ordination of the two perspectives. In fact, we claim that two students did successfully co-ordinate target and intention connections.

During the second phase of their exercise, the students constructed a tentative model for the connection of the data-centric and modelling perspectives on distribution (Fig. 5). They typically tended to gravitate towards simple causal explanations when they observed the modelling distribution, which, in some sense, generates the data. As soon as they discovered the random mechanism that underlies the generation of the throws, they sought to assimilate this unexpected obstacle encountered on the road to employing causal explanations. They substituted for true deterministic causality the quasi-causal actions of agents to operationalise randomness, in a parallel manner to how Piaget claims that mankind invents probability to operationalise randomness.

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

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

4 Second Study: TinkerPlots

The data used here come from a research study conducted in a rural secondary school. The data were collected during regular mathematics lessons designed by the researcher. In these lessons, the students spent extensive time working in pairs. The researcher interacted continuously with different pairs of students during the pair work phase in order to probe the reasoning and their understanding. The data collected included audio recordings of the lessons and video recordings of the screen output on the computer activity using Camtasia software. The researcher (Re) prompted students to use the mouse systematically to point to objects on the screen when they reasoned about computer-based phenomena in their attempt to explain their thinking.

At the first stage, the audio recordings were fully transcribed and screenshots were incorporated as necessary to make sense of the transcription. The most salient episodes of the activities were selected. The data were subsequently analysed using progressive focusing (Robson 1993).

This article focuses on one pair of students, George and Rafael. Insights similar to those reported below were evident in the analysis of the sessions of other pairs of students, however, George and Rafael provided (in the researcher's view) the clearest illustration of how students successfully build connections between probability, statistics, and simulated phenomena within a *TinkerPlots*2 computer-based environment.

The students spent three lessons watching instructional movies that show how to use *TinkerPlots2* features to build simulations and spent three lessons (40–45 minutes each) building simulations. In the fourth lesson, they watched a *TinkerPlots2* movie that showed how to use *TinkerPlots2* features to build a data factory that simulates real phenomena. Students were shown how to set a sampler as a mixer, or a spinner, or a histogram, or a density function.

Students were asked to use the tools of *TinkerPlots2*, in particular the "Data Factory" feature (see Fig. 6), to generate a number of individual "virtual students" to populate a "virtual secondary school", with each virtual student created using the student-defined probability distributions for each of the different variables (e.g. gender, name, height) necessary to identify one of the virtual students. As the simulation ran, the students observed the generation of data; and the distributions of the attributes of data.

Tinkerplots2 provided students with tools that exploit probability as a modelling tool using the sampler that is essentially a non-conventional form of probability distribution. The sampler gives students the option to set up the sampler as a spinner, in which they could visually assign different angles for the sectors that correspond to the probability of selecting a gender for each virtual student (Fig. 6).

5 Results

I joined George and Rafael as they began to use the tools of *TinkerPlots*2 to create a factory for their virtual secondary school. They began by creating a sampler to assign gender to each virtual student. To do this, they used a "spinner" (see Fig. 6), in which they could visually assign different angles to correspond to the probability of a given outcome. As can be seen in Fig. 6, they chose unequalised angles for the sectors, thus giving unequal probabilities of getting male (m) students and female (f) students; in this case there was a much larger probability of getting male students than females. Students explained their choice:

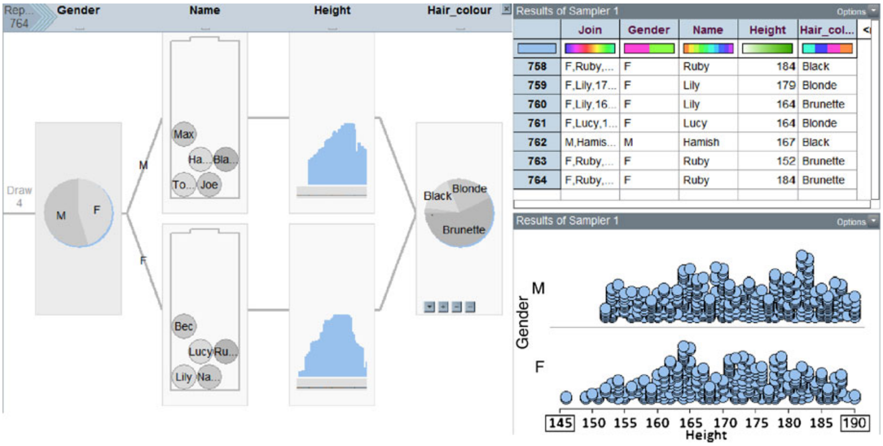


Fig. 6 Data factory that simulates a 'virtual school'

1. Ge: In schools there are generally more males.
2. Ra: Umm, generally the schools that I've been in there have been more males. This school, for example, has a lot more males than females.

After deciding on proportions of male and female students in the virtual school, George and Rafael chose to use the "mixer" function to give names to the virtual students. The boys created two different mixers, one for the names of each gender, and then decided what names to place in each mixer, from which one name was chosen at random for each virtual student of the appropriate gender. When the boys went on to give the virtual students another characteristic, height, Rafael quickly stated, "This is getting complex. For a student, we have a choice of gender, name, and height. We are going to have many data for each student". George agreed that although "each student is one, varied information is provided for a student".

Rafael and George seemed to find it complex when a person or an object is presented as comprising a set of attributes. It was interesting how the boys used a modelling approach as an intermediate step in attempting to perceive a holistic entity, such as a virtual student, as consisting of a cluster of pieces of data.

The students decided to introduce another attribute for the virtual students, height. The boys set up two samplers, one each for boys' and girls' heights, as a probability density function by drawing two curves as shown in Fig. 6. They also designated the additional attribute of hair colour and set up a sampler for this attribute as a spinner defining the sizes of its sectors.

Twenty minutes into the activity, the boys decided to have the data factory generate 675 virtual students. They ran the simulation to generate sample data for their virtual school and the researcher asked them to compare the distributions of height of the virtual students with the curves they drew in the sampler (see Fig. 7).

3. Ra: (pointing to the distribution on the bottom right of Fig. 7(a)). We had it rising there, then we had a drop, then we had a big rise there which is why I'm guessing all these came from before it dropped down there.

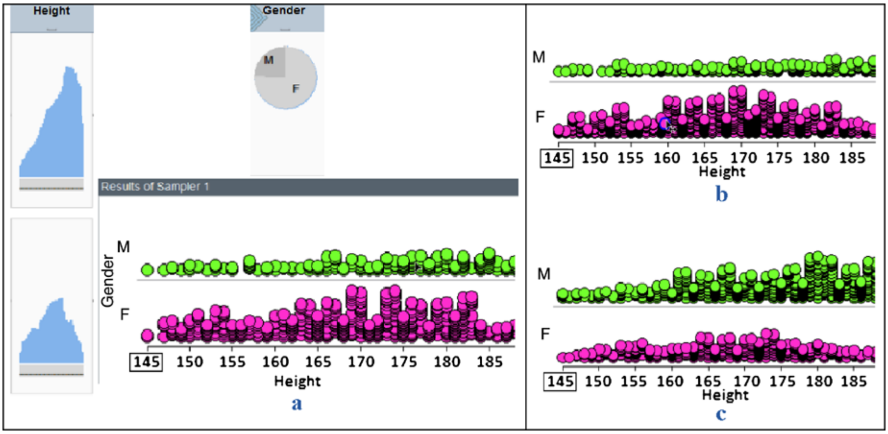


Fig. 7 Distributions created in the sampler and distributions of height

4. Ge: Well the females theory of the rise (pointing to the graph on the bottom left of Fig. 7(a)), which is here (pointing to 150–155 of the bottom left graph), and that goes down a bit before continuing, so it would be going, down, continuing up (170–175), around here (180–183), before decreasing again. With the males, we thought the height was a lot higher that was there (pointing to the graph on top left).
5. Ra: There, there's a lot less males though, so even, the end results aren't as packed there, the more spread out, or they seem to be spread out as much as the females but there's no piles. Like you can't see them as high as well as you can see the females height increases but you can actually see the increase there (175–185 right bottom).
6. Ge: There is an increase but it's not as prominent as the female increase.

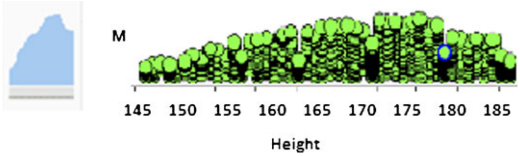
Rafael and George decided to make 1000 virtual students because they believed that the graphs would show the distribution of the height of males clearer. They observed:

7. Ra: But now you can see the increase in the males a lot better (pointing to the top Fig. 7(b)). Like you can see the towers higher, stacked higher than they were before. It's clear see that there is increase, but with the females is you can still see the increase, the huge increase the females have had as well.

Rafael paid attention to slices of prominent features of the distributions such as higher areas of accumulated data of the distribution. George attempted to equalise the number of male students with female students. After generating a new set of virtual students with a 50:50 ratio of male to female students, the boys observed the new distribution (Fig. 7(c)):

8. Ra: Yeah, it's a lot clearer to see the increase in the males (pointing to the distribution of male height), now because it is a probability it's not going to

Fig. 8 The distribution of male height they created in the sampler and the distribution of male height



look exactly like that (pointing to the distribution of male height they created in the sampler). There's going to be exemptions. But you can see, you can see the overall that it's increasing. Getting higher here before dropping down again (pointing to the distribution of male height), which is what our graph showed (pointing to the distribution of male height they created in the sampler). The females you could tell fairly compact in the middle … and there's not as much any to the side which is show here (pointing to the distribution of female height).

9. Re: Exemptions?[1]
10. Ra: I suppose there is always gonna be exceptions to the graph. They're not always gonna be exactly as we plan out. Because this is based on probability and probability, just because we see that (pointing to the distributions they created in the sampler), like it doesn't mean that it will follow that (100 %).

The above excerpt shows how Rafael's attention was, at this point, focused on the way the shape of the distributions of heights was changing compared to the distributions they created in the sampler. The extended discussion of the boys showed that Rafael could not recognise the absolute resemblance of the data distribution of heights with the distribution they created in the sampler due to uncertainty caused by probability.

In a following trial, George suggested creating a virtual school with 500, all-male students. To their surprise:

11. Ra: Well, it's steadily increasing with slight jumps. Like it lows there and jumps up a bit, which guessing that 140, 155 region there (pointing to the left graph in Fig. 8).
12. Ra: It's a 163 (left graph). And that 163 just there (right graph).
13. Ge: It's going up and just going down a little bit (right graph).
14. Ra: That's probably where that is come from 160. 164 here has just a low. But then it continues. Then it jumps back up and keeps going before dropping down a bit again here (pointing to both graphs at the same time).
15. Ge: Just there that one, for reason it's just steadily dropping (pointing between 180–185 of right graph).
16. Ra: … except there (180–185). That seems to have a sharp drop which I'm guessing is just off one of these areas here. Where it just seems to drop down (both graphs).

[1]When Rafael was asked about the word "exemptions", he shifted to using the word "exceptions".

17. Re: Do you believe that the final graph resembles of what you created in the sampler?
18. Ra: Yeah, fairly accurately.

After excluding female students from their model Rafael was able to see that the two distributions eventually resembled each other.

We saw how Rafael and George moved over to the modelling approach perceiving a holistic entity such as a student as consisting of a cluster of pieces of data having attributes such as gender, name, etc. When the students attempted to simulate a virtual school, through the activity of setting up the sampler as a spinner defining the sizes of the sectors they connected probability to simulated phenomenon. Similarly, when students set the sampler as a mixer and selected a sample randomly, again they drew connections between probability and simulated phenomena. When they set a sampler as a probability density function by drawing a curve based on the users' personal experiences, they drew connections between probability, statistics, and the simulated phenomenon.

When the boys were challenged to draw a curve that mirrored a probability density function of heights, they first identified the area where the most common heights were accumulated (what we could consider the "signal", though they did not use this term) and then talked about the variation (what we could consider the "noise") of heights around that area. These students seemed to realise that the nature of a reasonable approximation of real or simulated phenomena lies in the relationship between signal and noise. A good co-ordination of signal and noise requires a fairly good knowledge of the phenomenon at hand, so that the students could design the *TinkerPlots* Data Factory in such a way that it would generate sample data which would resemble as much as possible the real-world phenomena it was intended to model. These boys seemed to realise the importance of the choices they made as modellers, thus they focused their attention on the distribution of heights they created in the sampler and the other attributes, such as gender and hair colour. As we witnessed from the students' activity, the positioning of probability as a modelling tool to build models in computer-based simulations brings distribution of data (e.g. distribution of students' heights) into the foreground, not as a pre-determined entity, but as a non-fixed entity, open to debate.

When boys were working on the task of comparing the data distribution of the simulated heights to the distribution they created in the sampler, they appeared to have some difficulty understanding the nuances of probability. Nonetheless, they were not without insight. The students articulated situated heuristics, for example, "there is always gonna be exceptions to the graph. They're not always gonna be exactly as we plan out. Because this is based on probability and probability, just because we see that (pointing to the distributions they created in the sampler), like it doesn't mean that it will follow that (100 %)". This situated abstraction can be interpreted as a construction of a relatively naïve conception of the use of probability both as a modelling tool and as a measure of confidence that one can give to the model they created. However, another possible interpretation of this situation could be that when the model is run a few times, there is stability in the peaked data but there is some variation observed in the general details of the shape, thus Rafael could not recognise or accept the absolute resemblance of the distributions of heights to the curves students created in the sampler due to the existence of many variables that caused variation. When they excluded one variable from the attributes, the students were better able to draw conclusions about the resemblance of the distribution of the generated data with what they designed in the sampler. By excluding variables from the model, thus reducing complexity, the students expressed an approach to introducing the variables in a systematic way that might benefit students exploring the connections between probability and the statistical data generated by simulated phenomena.

6 Discussion

The use of probability as a modelling tool was presented in the two simulation activities. In the activities, students had direct control over the modelling perspective. In the *BasketBall* simulation exercises, students could determine the position of the arrows (spread) and the handle on the slider (spread), or they could set values for individual events, hence attributing probabilities to a range of possible outcomes in the sample; whereas in the virtual school exercise, the students were able to draw a curve to define a probability density function and set up a sampler as a spinner to define the probability of an event—and through this direct control of the modelling perspective, they could influence indirectly the data generated and thus the data-centric perspective.

6.1 Student Responses

Using the *Basketball* simulation, the students constructed a bi-directional connection of the data-centric and modelling perspectives on distribution by associating controls in the simulation interface that represent parameters of the probability distribution to the shape and position of histograms of data. Students who used the *BasketBall* simulation constructed three interpretations for bridging the modelling and the data-centric perspectives on distribution:

General Intention (I_G), Stochastic Intention (I_{st}) and Target (T).

Similarly, the *Tinkerplots2* exercises provided different options of using samplers that are non-conventional representations of a probability distribution to describe variation of data in real or imagined phenomena. In the virtual school simulation, when students attempted to associate the data distribution of the simulated heights to the curve that mirrored a probability density function of heights, they could not recognise the absolute resemblance of the distributions of heights to the curves that the students created in the sampler.

There appear to be three possible causes for this difference: (i) messiness caused by variation among many variables; (ii) relationship between signal and noise; and (iii) knowledge of events/situations to be simulated.

(i) *Messiness caused by variation among many variables*. In the *BasketBall* simulation experiments, students interacted with one variable at a time, a simple or "clean" variation (and incorporated variation into the other variables in a consistent way); whereas in the *Tinkerplots*2 experiment, the data factory generated data from many different variables at once, creating a "messy" variation. The messiness seemed to inhibit students' ability to draw useful conclusions.

Variation was not controlled in the context of virtual school simulation. When the students excluded one variable from the attributes, they were better able to draw conclusions about the resemblance of the distribution of the generated data with what they designed in the sampler. By excluding variables from the model, thus reducing complexity, the students expressed an approach to introducing the variables in a systematic way that might benefit students exploring the connections between probability and the statistical data generated by simulated phenomena.

In the *BasketBall* simulation, variation in the trajectories of the balls was handled in a consistent way. Variables such as shooting position, speed and angle had all been fixed by default but with the option of implementing variation, which itself can be increased or decreased in size. The task directed the attention of the students to causality (speed and angle of throw), since variables without any arrows on them were entirely determined, never varying from throw to throw. The student could begin to let go of determinism (Prodromou 2008; Prodromou and Pratt 2006, 2009) by implementing variation in the determining variables, perhaps perceived as allowing for skill level. This was a fairly natural step since it felt inappropriate that once a successful throw was discovered, the thrower would succeed every time.

Thus, a student was allowed to choose to have only one variable with variation and s/he might choose to gradually increase that variation before introducing a second variable with variation. I believed that, with this additional control over variation, students were able to connect the deterministic to the stochastic in a more fine-grained way.

(ii) *Relationship between signal and noise*. A reasonable approximation of the real or simulated phenomena being modelled depends exclusively on how the students express the relationship between signal and noise. Related to the question of messiness, it was observed that having an interface that provided students with manipulable objects—signal and noise—as instantiated in the form of quasi-concrete objects, facilitated the students' ability to appreciate better the relation between signal and noise.

Where the students in the *BasketBall* simulation could manipulate both signal and noise of a single variable's model, and data output, the virtual school simulation study did not reveal this distinction as clearly.

In the *BasketBall* simulation, the students were able to transform the *modelling* distribution directly, and this direct control over the variables controlling the computer's generation of data allowed indirect control over data generated, and thus over what they saw in the data-centric perspective on distribution.

The students interacted at the top-level by moving either the arrows or the handle on the slider of a variable (Fig. 3) and observed the impact of their actions on the graphical representation of the modelling distribution. The students were able to connect those features (handle and distance between the arrows) with notions of average and spread and the shape of distribution. Moreover, the students were able to transform the modelling perspective on distribution by attributing frequencies to the range of possible outcomes in the sample, creating therefore a model out of which the data-centric perspective would (Fig. 4).

In the *Tinkerplots*2 study, the students set up the sampler as a probability density function by drawing a curve. The students simply identified the area where the most common heights were accumulated (signal) and then talked about the variation (noise) of heights around that area. The students working on these exercises seemed to lack any hands-on activity that would support their comprehension of the relationship between signal and noise in their simulated event.

(iii) *Knowledge of events/situations to be simulated*. Along the same lines as the previous two points, students found it easier to reason about phenomenon about which they had some knowledge. In the *BasketBall* simulation, students were able to draw more useful conclusions than in the virtual school simulation study, where they were reasoning about a complex situation, which resembled their own real-world school environment in some ways, and thus their own school was used as a model for reasoning, and the determination of the various variables of their school distracted them. A good co-ordination of signal and noise requires a fairly good knowledge of the phenomenon at hand, so that modellers could make "optimal" choices when designing the model; which would generate sample data that resembles as much as possible the real-world phenomena it is intended to model. The familiarity and simplicity of the playful *BasketBall* simulation enhanced students' ability to make easy connections between aspects of the game and the statistical/probabilistic models related to the two perspectives on distribution. By contrast, the students struggled while they attempted to model the attributes (variables) from their virtual school, which was more complex in ways that they had not previously considered. Their knowledge of the complex cultural context of the event being modelled had cognitive implications on students.

6.2 Implications

Teaching probability at Year 10 requires students to assess likelihood and assign probabilities using experimental and theoretical approaches (ACARA 2010). The Mathematics K-10 Syllabus in Australia has focused on "the calculation of theoretical and experimental probabilities of simple and compound events". The Australian Curriculum and Reporting Authority advocates the broadening of probability and statistics in the school curriculum: "Statistics and probability initially develop in parallel and curriculum then progressively builds the links between them" (ACARA 2010, p. 2).

Students need to be able to make straightforward mappings of aspects of a real world situation onto probabilistic structures. By reference to students' existing knowledge of real events, a teacher could encourage students to link the behaviour of a real world situation to a probabilistic structure, analysing the specific characteristics of the real world situation being modelled and the factors that could affect its behaviour in response to possible variations.

A simple example is the following. Suppose a teacher could use various simulation activities that will help students to build a model of a throw of a ball as a starting point to model a real event and considers many questions, most likely based on questions about playing traditional sports with a ball. The questions could address both variables that affect the throw of the basketball, such as release angle, speed, height and distance to the basket, and characteristics of the basketball player are appropriate to reflect features of real-world games. Questions to ask might include:

- What are the variables that affect the throw of the ball?
- Can we change those attributes?
- Do they affect each other?
- What are the characteristics of the skill of a basketball player?

The answers to such questions will take into account the factors that impact the trajectories of the balls, so the trajectories of the balls will be viewed as consisting of a set of data. In such a way, the desired aim of the activities is to informally introduce students to the independent variables, representing the value being changed and the dependent variable that is the outcome of the changing independent variables. This latter distinction between independent versus dependent variables will give rise to a discussion about the mechanism underlying the simulation and determines the trajectories of the balls. That mechanism, as in the BasketBall simulation, could either be strictly deterministic based on a mathematical algebraic formulae derived from Newton's laws of motion; or a stochastic model based around a probability distribution that allowed variation in the trajectories. Indeed, "chance variation rather than deterministic causation explains many aspects of the world" (Moore 1990, p. 99).

The purpose of probability and statistics is to explain variation in meaningful ways using models. Probability distributions exist, for example, to explain the theoretical random variation that occurs when measurement of a constant quantity is carried out many times or many trials of generators are performed many times. This suggests that students, at the outset, need to be provided with tasks in which they can investigate repeated measurements of a constant quantity, thus observing that such measurements are subject to random variation. Students should be challenged to measure a specific object repeatedly and observe that although the weight of a specific object is constant, they are apt to get slightly different values. For the modelling task of this activity, it is better to encourage students to formulate the idea of the specific object having a typical weight. Comparing different measurements of objects leads to new insights into the relationship between a concrete situation and its probabilistic modelling. After making comparisons between all the available repeated measurements, the most common weight of the specific object would be chosen by students from the data. Considering this "typical weight of the specific object" resulted from the comparable data, the actual weight of the object might be modelled as:

[Typical weight of the specific object-constant value, typical weight of the specific object + constant value]

Such an expression can be translated as:

[signal − noise, signal + noise]

The way students express this relationship between signal and noise is of paramount importance while they are working at the concrete level (Henry 2001; Batanero et al. 2005) observing a real situation. To help students move from developing models, expressions focusing on the relation of signal and noise involve elements of abstraction and simplification of reality with respect to the real situation studied. This might help students move to the second modelling stage (Batanero et al. 2005) and develop models that are represented in a symbolic system suitable for probability calculus.

There is also the issue in statistics of unexpected variation or "common cause" variation, that is, variation that it is difficult to be assigned to any particular source and does not fit any particular distribution or can be expressed as related to noise. Unexpected variation may be caused by non-random behaviour, or perhaps variation as related to a combination of variables; causing messiness caused by variation among many variables.

Students should be taught from their early years in primary school years the importance of relating variation to noise when modelling real situations, so they would develop a view of variation around a signal and variables as the very reason for the existence of the discipline of statistics (Shaughnessy and Ciancetta 2001) because "thinking about variability is the main message of statistics" (Smith 1999, p. 249).

It is important, however, that such discussions not be diminished throughout the years of schooling. Notwithstanding, when students study formal statistics courses, they will use confidence intervals or hypothesis testing to promote decision-making about the variation in problem-solving situations and whether a particular conclusion is satisfactory or not. Such formal procedures exclusively rely on rules tightly related to probability. Teachers and researchers in probabilistic cognition nevertheless observed that usually students at the beginning of their formal statistical courses rely heavily on rules and they fail to consider the connections between probability and the data generated by real or imagined simulated phenomena. It is therefore important, from the middle years of schooling, to explore ways to draw connections between probability and data.

Students need to be able to have a good understanding of the Big Ideas listed above including foundational concepts such as variation and signal, variables that should be considered in ways that acknowledge students' starting points and build increasingly complex structures for understanding models. Activities should include many different models and these should be specifically linked to each other where appropriate. As well as to the underlying uncertainty whether the data that is being generated from some assumed theoretical distribution sufficiently imitates the fundamental features of the process that generated the data.

Computer simulations such as the *BasketBall* simulation or *Tinkerplots2* are not the end of considering probability as a modelling tool in the curriculum. These must lead to consideration of the cognitive importance of associating a probability distribution with a random experiment, and with the important part played by building links between variation, theoretical models, simulations, and probability. These are the areas where sophisticated understanding and application of chance can be useful to students in decision making and modelling when modelling everyday phenomena.

References

Australian Curriculum, Assessment and Reporting Authority (2010). Australian Curriculum: Mathematics. Version 1.1. Online: http://www.acara.edu.au.

Batanero, C., Henry, M., & Parzysz, B. (2005). The nature of chance and probability. In G. A. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 15–37). New York: Springer.

Borovcnik, M. (2011). Strengthening the role of probability within statistics curricula. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics—challenges for teaching and teacher education, a joint ICMI/IASE study: the 18th ICMI study* (pp. 71–83). New York: Springer.

Chaput, M., Girard, J.-C., & Henry, M. (2011). Frequentist approach: modelling and simulation in statistics and probability teaching. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics—challenges for teaching and teacher education, a joint ICMI/IASE study: the 18th ICMI study* (pp. 85–95). New York: Springer.

De Moivre, A. (1967/1756). *The doctrine of chances: or a method of calculating the probabilities of events in play*. New York: Chelsea (original work published in 1756, London: A. Millar).

Henry, M. (2001). *Autour de la modélisation en probabilités [Around modelling in probability]*. Besancon: Presses Universitaires Franc-Comtoises.

Hoerl, R. W., & Snee, R. D. (2001). *Statistical thinking: improving business performance*. Pacific Grove: Duxbury.

Laplace, P. S. (1995/1814). *Théorie analytique des probabilités [Analytical theory of probabilities]*. Paris: Jacques Gabay (original work published 1814).

Makar, K., Bakker, A., & Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning*, *13*(1), 152–173.

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, *8*(1), 82–105. Online: http://www.stat.auckland.ac.nz/~iase/serj/SERJ8(1)_Makar_Rubin.htm.

von Mises, R. (1952/1928). *Probability, statistics and truth (J. Neyman, O. Scholl, & E. Rabinovitch, Trans.)*. London: William Hodge and Company (original work published 1928).

Moore, D. S. (1990). Uncertainty. In L. S. Steen (Ed.), *On the shoulders of giants: new approaches to numeracy* (pp. 95–137). Washington: National Academic Press.

Pratt, D. (2011). *Re-connecting probability and reasoning about data in secondary school teaching*. Paper presented at 58th ISI World Statistics Congress, Dublin, Ireland. Online: http://isi2011.congressplanner.eu/pdfs/450478.pdf.

Prodromou, T. (2007). Making connections between the two perspectives on distribution. In D. Pitta-Pantazi & G. Philippou (Eds.), *Proceedings of the fifth congress of the European*

Society for Research in Mathematics Education, Larnaca, Cyprus (pp. 801–810). Online: http://ermeweb.free.fr/CERME5b.

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

Prodromou, T. (2012). Students' construction of meanings about the co-ordination of the two epistemological perspectives on distribution. *International Journal of Statistics and Probability*, *1*(2), 283–300. doi:10.5539/ijsp.v1n2pxx.

Prodromou, T., & Pratt, D. (2006). The role of causality in the co-ordination of the two perspectives on distribution within a virtual simulation. *Statistics Education Research Journal*, *5*(2), 69–88. Online: http://www.stat.auckland.ac.nz/~iase/serj/SERJ5%282%29_Prod_Pratt.pdf.

Prodromou, T., & Pratt, D. (2009). Student's causal explanations for distribution. In V. Durand-Guerrier, S. Soury-Lavergne, & F. Arzarello (Eds.), *Proceedings of the sixth conference of the European Society for Research in Mathematics Education*, Lyon, France (pp. 394–403).

Renyi, A. (1992/1966). *Calcul des probabilitiés [Probability calculus] (L. Félix, Trans.)*. Paris: Jacques Gabay (original work published 1966).

Robson, C. (1993). *Real world research*. Oxford: Blackwell.

Shaughnessy, J. M., & Ciancetta, M. (2001). Conflict between students' personal theories and actual data: the spectre of variation. In *Second round table conference on research on statistics teaching and learning*, Armidale, New South Wales, Australia.

Smith, T. M. F. (1999). Discussion. *International Statistical Review*, *67*(3), 248–250.

Sorto, M. A. (2006). Identifying content knowledge for teaching statistics. In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education. Proceedings of the seventh international conference on teaching statistics*, Salvador, Brazil. Voorburg: International Statistical Institute [CDROM]. Online: http://www.stat.auckland.ac.nz/~iase/publications/17/C130.pdf.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading: Addison-Wesley.

Wild, C., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*, 223–266.

Promoting Statistical Literacy Through Data Modelling in the Early School Years

Lyn D. English

Abstract This chapter addresses data modelling as a means of promoting statistical literacy in the early grades. Consideration is first given to the importance of increasing young children's exposure to statistical reasoning experiences and how data modelling can be a rich means of doing so. Selected components of data modelling are then reviewed, followed by a report on some findings from the third-year of a three-year longitudinal study across grades one through three.

1 Statistical Literacy and Young Children

Across all walks of life, the need to understand and to apply statistical literacy is paramount. Statistics underlie not only every economic report and census, but also every clinical trial and opinion poll in modern society. Our unprecedented access to a vast array of numerical information means we can engage increasingly in democratic discourse and public decision-making—that is, provided we have an appropriate understanding of statistics and statistical literacy. Statistical literacy, however, requires a long time to develop and must begin in the earliest years of schooling (English 2010; Franklin and Garfield 2006; Watson 2006). Numerous definitions of statistical literacy abound (e.g. Gal 2002). The notion adopted here is that of Watson (2006), namely, statistical literacy is "the meeting point" of the statistics and probability strand of a given curriculum and "the everyday world, where encounters involve unrehearsed contexts and spontaneous decision-making based on the ability to apply statistical tools, general contextual knowledge, and critical literacy skills" (Watson 2006, p. 11).

Young children are very much a part of our data-driven society. They have early access to computer technology, the source of our information explosion. They have daily exposure to the mass media where various displays of data and related reports (e.g. weather reports and popularity contests) can easily mystify or misinform, rather than inform, their young minds. The need to advance children's statistical

L.D. English (✉)
Queensland University of Technology, Brisbane, Queensland, Australia
e-mail: l.english@qut.edu.au

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_23,

reasoning abilities, from the earliest years of schooling, has thus been stressed in recent years (e.g. Langrall et al. 2008; Shaughnessy 2010; Whitin and Whitin 2011). One approach to enhancing children's statistical abilities is through data modelling (English 2010; Lehrer and Romberg 1996; Lehrer and Schauble 2007).

2 Data Modelling

Data modelling is a developmental process, beginning with young children's inquiries and investigations of meaningful phenomena, progressing to identifying various attributes of the phenomena, and then moving towards organising, structuring, visualising, and representing data (Lehrer and Lesh 2003). Data modelling should be a fundamental component of early childhood curricula, yet there exists limited research on such modelling and how it can be fostered in the early school years. The majority of the research has been concerned with the secondary and tertiary levels, with the assumption that primary school children are unable to develop their own models and sense-making systems for dealing with complex situations (Greer et al. 2007).

Recent research, however, has indicated that young children do possess many conceptual resources which, with appropriately designed and implemented learning experiences, can be bootstrapped toward sophisticated forms of reasoning not typically seen in the early grades (e.g. Clarke et al. 2006; Clements et al. 2011; English and Watters 2005; Papic et al. 2011; Perry and Dockett 2008). Most research on early mathematics and science learning has been restricted to an analysis of children's actual developmental level, which has failed to illuminate their potential for learning under stimulating conditions that challenge their thinking (Ginsburg et al. 2006; Perry and Dockett 2008). Data modelling provides rich opportunities to advance young children's statistical development and reveal their capabilities in dealing with challenging tasks.

I now address two core components of data modelling that are pertinent to the activities addressed in the research reported here, namely, structuring and representing data. I also explore the metarepresentational competence that children display in data modelling.

3 Structuring and Representing Data

Models are typically conveyed as systems of representation, where structuring and displaying data are fundamental; the structure is constructed, not inherent (Lehrer and Schauble 2007). However, as Lehrer and Schauble indicated, children often have difficulties in imposing structure consistently and often overlook important information that needs to be included in their representations or alternatively, they include redundant information. Providing opportunities for young children to structure and display data in ways that they choose, and to analyse and assess their representations is important in addressing these early difficulties. Yet young children's typical exposure to data structure and displays has been through conventional instruction on standard forms of data representation. The words of Russell (1991) are still timely today:

> We have two choices in undertaking data analysis work with students: we can lead them to organising and representing their data in a way that makes sense to us, or we can support them as they organise and represent their data in a way which makes sense to them. In the first case, they learn some rules—and they learn to second-guess what they are supposed to do. In the second case, they learn to think about their data. Students need to construct their own representations and their own ways of understanding, even when their decisions do not seem correct to adults (p. 160).

As children construct and display data models, they generate their own forms of inscription. By the first grade, children already have developed a wide repertoire of inscriptions, including common drawings, letters, numerical symbols, and other referents. As they invent and use their own inscriptions, children also develop an "emerging meta-knowledge about inscriptions" (Lehrer and Lesh 2003). Children's developing inscriptional capacities provide a basis for their mathematical activity. Indeed, inscriptions are mediators of mathematical learning and reasoning; they not only communicate children's mathematical thinking but they also shape it (Lehrer and Lesh 2003; Olson 1994). As Lehrer and Schauble (2006) stressed, developing a repertoire of inscriptions, appreciating their qualities and use, revising and manipulating invented inscriptions and representations, and using these to explain or persuade others, are essential for data modelling.

4 Metarepresentational Competence

Children's use of inscriptions plays an important role in their development of metarepresentational competence (diSessa 2004; diSessa et al. 1991). Such competence includes students' abilities to invent or design a variety of new representations, explain their creations, understand the role they play, and critique and compare the adequacy of representations. The use of "meta" to describe these capabilities, as diSessa's emphasises, is to indicate that no specific representational skills are implied. Unlike the standard representational techniques students might have learned from specific instruction, metarepresentational competence encompasses students' "native capacities" (diSessa 2004, p. 294) to create and re-create their own forms of representation. Their skills here are more "broadly applicable, more flexible and fluid" (diSessa et al. 1991, p. 118), and are not just confined to a narrow set of instructed representations. In essence, diSessa (diSessa and Sherin 2000) and his colleagues coined the term metarepresentational to "describe the full range of capabilities that students (and others) have concerning the construction and use of external representations" (p. 386). Indeed, diSessa's research has shown that students do possess a "deep, rich, and generative" understanding of representations, which seems to exist before instruction and is independent of it (p. 387). It appears students are particularly strong at inventing and modifying representations, which diSessa (2004) refers to as "hyper-richness."

Another issue that has received limited attention with respect to children's metarepresentational competence is the joint development of metarepresentational and conceptual competence (diSessa 2004). As diSessa noted, research is limited here and the role of student-created representations in conceptual development is rather complex. Questions needing attention include how certain strengths or limits of metarepresentational competence might advance or hinder conceptual competence, and whether metarepresentational competence and conceptual competence develop jointly.

In the remainder of this chapter, I report on some findings from the third year of a three-year longitudinal study that implemented data modelling across grades one to three. Specifically, I consider how the children structured and represented data they had collected, and the conceptual and metarepresentational competence they displayed in doing so.

5 Methodology

5.1 Participants

The participants were from an inner-city Australian school. In the first year of the study, three classes of first-grade children (2009, mean age of 6 years 8 months) and their teachers participated. The classes continued into the second year of the study (mean age of 7 years 10 months), and finally, two classes continued into the third year (mean age of 8 years 8 months, $n = 39$). A seventh-grade class (age range of 12–13 years) also participated in one of the activities during the second year. For all activities, the children worked in small groups.

5.2 Design

A teaching experiment involving multilevel collaboration (at the level of student, teacher, and researcher) was adopted throughout the study, where the developing knowledge of all participants was the focus (English 2003; Lesh and Kelly 2000). Such an approach is concerned with the design and implementation of experiences that maximise learning at each level. The teachers' involvement in the research was vital; hence regular professional development meetings were conducted where we planned activities, reflected on the implementation of the activities in preparation for the next activity, and reviewed the children's learning. This chapter addresses aspects of the student level of development.

5.3 Procedures

All activities were implemented by the classroom teacher, while a senior research assistant and I observed the children and facilitated group progress. No direct instruction, however, was given by either the teachers or the researchers. The children were encouraged to work as an independent group and to generate their own ways of working statistically. The children had been exposed to minimal statistical experiences in their school curriculum and most of the representations they created, as indicated later, had not been taught.

The data addressed in this chapter are drawn from the last two of three activities implemented in the final year of the study. The three activities engaged children in posing and refining questions; identifying, deciding on, and measuring attributes; developing and conducting a survey; collecting and recording data; organising, interpreting, analysing and representing data; and developing data-based explanations, arguments, inferences and predictions.

Literature, both purposefully created and commercially available, was used as a basis for the problem context in each of the activities implemented across the three years. It is well documented that storytelling provides an effective context for mathematical learning, with children being more motivated to engage in mathematical activities and displaying gains in achievement (van den Heuvel-Panhuizen and Van den Boogaard 2008).

For the second activity, *Investigating and Planning Playgrounds*, the story book *Hot Cha Cha* (Nobisso 1998) was initially read to the children. The activity was then introduced, with the children initially posing questions that might help them find out more about their classmates' thoughts on their new playground. In their groups, the children then created four survey questions and were to provide four answer options for each question (e.g. one group posed the question, "How long do you spend on each piece of equipment?" with the response options of 30, 15, 5, and 20 minutes). On answering their own questions, each group chose one focus question to which the other groups were to respond. The children were to initially predict how their focus question might be answered by the remaining groups.

Each group subsequently analysed all their collected data for their focus question and were to display their findings using their choice of representation. The children were encouraged to structure and represent their findings in more than one way, with no specific direction given on how they might do so. They were supplied with a range of recording material including blank chart paper, 2.5 cm squared grid paper, chart paper displaying a circle shape, and chart paper displaying horizontal and vertical axes (without marked scales). The children could use whichever of these materials they liked; no encouragement was given to use any specific recording material. The children were also given recording booklets. On completion of the activity, the groups reported back to their class peers on their final data models.

For the third activity, the story book *Simon and Catapult Man's Perilous Playground Adventure* (Smiley 2009) was used as an introduction. A class discussion was then held on the different features the children recalled of their new playground, with students asked to suggest ways of categorising these features. Next, the students explored their playground to find and record more specific data on various aspects, such as the number of foot/hand holds in the rock climbing wall, the number of triangle frames for the flying fox, and so on. The children did not take any measuring instruments with them; rather, they spontaneously used informal units of measure such as their estimated length of their stride to determine the perimeter of the playground.

On their return to the classroom, a brief class discussion was conducted on what the children had discovered. In their groups, the children were to then identify key features they would like in their desired playground and record and classify these. Once the children had recorded their "wish list" of playground features, they were to choose ways to represent one category of their desired playground data and show on their representations how their desired playground category differed from the existing playground.

As in the previous activity, the children were provided with a range of materials on which to record and represent their data. As before, no directions were given on material selection and use. Following the completion of their representations, the children were to record their answers to the questions, "Is the Central Play Space different to the playground you would like to design?" and "How does your representation show the difference?" Finally, the children reported back to the class on their desired playground and how their representation displayed a comparison between the desired and existing playgrounds.

5.4 Data Collection and Analysis

In each of the two third-grade classrooms, two focus groups (of mixed achievement levels and chosen by the teachers) were videotaped and audiotaped for both activities. Altogether there were nine groups of children, five in one class and four in the other, who completed the activities. All artefacts were collected and analysed, along with the transcripts from the video and audio tapes. Iterative refinement cycles for analysis of children's learning (Lesh and Lehrer 2000) were used, together with constant comparative strategies (Strauss and Corbin 1990) in which data were coded and examined for patterns and trends.

For both activities, the analysis of the children's transcripts and artefacts focused on the conceptual resources (diSessa 2004) the children used, with a focus on: (a) the ways in which the children structured and represented their data; (b) the "completeness (shows all relevant information)," and "compactness (better use of space)" (diSessa 2004, p. 313) of their representations; and (c) evidence of conceptual and metarepresentational competence and any possible links between these.

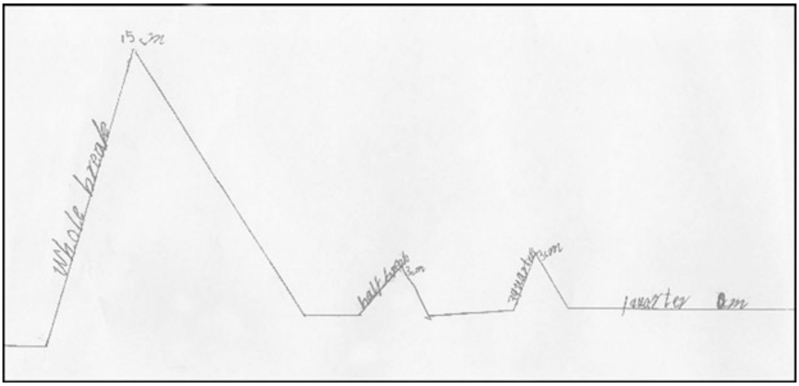


Fig. 1 'Heart Monitor' Graph

6 Selected Findings

6.1 How Children Structured and Represented Their Data

For activity 2, of the nine student groups, seven created two or more representations, with one group creating four representations. All but one of the student groups created more than one representation for activity 3, with five groups constructing three or more representations.

Vertical bar graphs and circle graphs were the most popular forms of representation for both activities, with the latter especially the case for activity 2, despite the children not having been taught circle graphs. In activity 2, seven groups created a vertical bar graph using the 2.5 cm squared grid paper and seven, a circle graph, with six groups creating both forms of representation. Beginning with a blank sheet of paper, two groups created a vertical bar graph and one group, a line graph. One group made use of a list of their response options and used tally marks to collate their data prior to representing it.

Two groups chose unusual representations for activity 2, such as the "heart monitor" representation created by James' group (Fig. 1). As James explained, "Okay, how are we going to show it our own way?! Let's think, just think. I know, like at the hospital, so like… up here (indicates a line graph similar to a heart rate monitor display). So first we're going to maybe do about 3 cm" (he places his ruler horizontally across the middle of the paper).

Another group member further added, "We're doing like a doctor's sort of thing how they go like that (indicating a rise and fall) … like a doctor does… yes, how they have those lines, so we're sort of doing it with maths though." This was an interesting and creative representation, which seemingly held meaning for the children as a way to represent their data, supporting diSessa's (2004) findings that "students are strikingly good at certain metarepresentational tasks, especially designing representations (p. 298)". Furthermore, this representation demonstrated the children's use of analogical reasoning (English 2004) as they identified a relationship between a heart monitor graph and a means of displaying the data they had gathered.

For activity 3, bar graphs using the 2.5 cm squared paper dominated the representations, with all but one group choosing this material. One group also chose to create their own grid paper representation. There were multiple uses of bar graphs, the increase due in large part to the task design involving a greater number of attributes to consider. Four of the nine student groups created other representations along with their bar graphs, which included circle graphs (two groups), partial line graphs (two groups), and lists and written explanations (four groups). The group that created the "heart monitor" graph in the previous activity decided to create two more of this form, in addition to two other representations. The first "heart monitor" representation only showed the desired improvements for the rock wall ("6 different shapes," "11 more rocks," and "3 metres higher"), with the children ruling "up and down" lines that were measured in centimetres corresponding to the quantities desired (e.g. 6 cm for "6 different shapes"). As one child explained, "I've done 10 cm because I couldn't use a metre ruler so I did 10 cm up and I put it as 10 m long." The second representation comprised two graphs of this nature, one displaying the "old" and "new" features for "different shapes" and the other, "old" and "new" heights. As another child explained, "Well, we did the same like that except we did 6 cm for 6 different shapes of the rocks." As indicated later, the children displayed metarepresentational competence in deciding whether this representation would be suitable for this third activity. Children's construction of varied representations in both activities and their abilities to explain and justify their constructions reflect their "native capacities" (diSessa 2004) to create and re-create their own forms of representation.

6.2 Completeness and Compactness of the Representations

The conceptual resources children displayed in developing their representations included their use of inscriptions and colour in all their representations and their use of space in their bar graph constructions (diSessa 2004). For activity 2, of the 10 bar graphs created (one group completed two), seven displayed inscriptions on both the vertical and horizontal axes, while the remainder omitted numbering the vertical axis, relying instead on a visual counting of the squares on their grid paper. Six bar graphs featured spacing between the drawn bars, presumably to facilitate reading of the graph, with one group commenting, "I'll leave a space so it is neat."

Interestingly, two groups who used the grid paper in a landscape position realised they did not have sufficient squares to represent the data for a couple of their response options and hence chose to colour two columns for the one response. All bar and line graphs featured colour, with one group debating whether, in creating a new representation, the same colours should be used ("We colour it exactly the same, and then we're going to colour code exactly as we did on that" [bar graph]; another group member disagreed, however, stating, "We don't need to.") Children's debate over the use of colour could be considered indicative of further development of metarepresentational competence involving a critical assessment of re-representations.

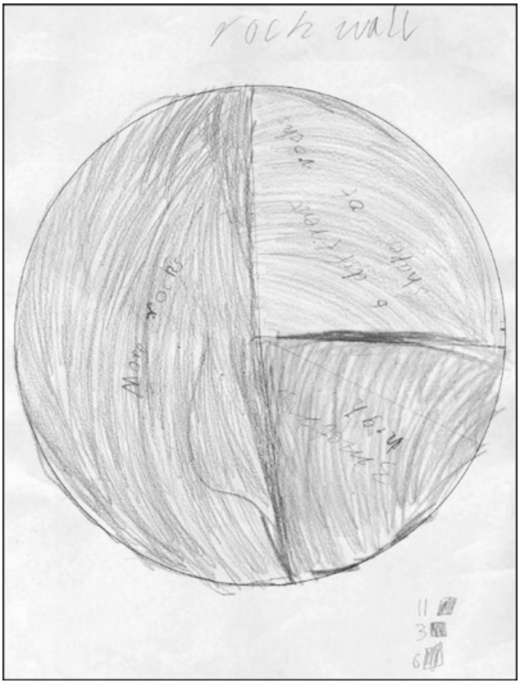


Fig. 2 Example of a Circle Graph

For those groups that created circle graphs for activity 2, all used inscriptions to indicate the proportion of each response option. These inscriptions included labelling of the option or using a colour key (two instances of the latter), as well as recording the number of each response and/or recording a percentage (five instances of the latter). The children's use of percent and percentage was an interesting finding, as explained in the next section.

The increase in representations created for activity 3 was accompanied by a broader and more complex set of inscriptions as children displayed comparisons between existing and desired playground features. Of the 14 bar graphs created across all the groups, only three did not record inscriptions on both axes; for two of the groups that did not number the vertical axis, but labelled the horizontal axis, totals were recorded at the top of the bars. The two circle graphs comprised inscriptions, such as one graph that divided the sectors into overarching categories such as "rock wall" with the subcategories of "old: 19 rocks, not gripy, small" and "new: 35 rocks, gripy, tall" inscribed in the sectors.

The other circle graph (titled, "rock wall," Fig. 2) was divided into three sectors, displaying the inscriptions, "more rocks," "3 metres high," and "6 different shapes of rocks."

Notice how the children made an effort to apportion the sectors according to the numerical value being represented, as indicated on the key the children created. This group also generated further representations showing the differences in existing and

Table 1 An example of a group's comparative list

What we want	What we have
We want the spider web two metres taller and tighter	1 × spider web
We want the chain climbing frame to be $1\frac{1}{2}$ metres taller	1 × chain climbing frame
We want the rockwall to be $1\frac{1}{2}$ metres taller and 35 grips with more grip	1 × rockwall
We want the 2 ladders to be $1\frac{1}{2}$ metres taller	2 × ladders
We want the fort to be $1\frac{1}{2}$ metres taller	1 × fort

desired playground features, including the "heart monitor" representations that they used in the previous activity (as discussed further in the next section).

One group, who created three different representations, included a partial line graph where they recorded at the top of their graph, "% we changed the climbing frame, chain links, rock wall, and ladder" labelling their graph with 7 %, 13 %, 3 %, and 0 %, respectively. This reflected the children's fascination with percent and percentage as indicated in the next section.

Given that children's written language is a powerful "constructive resource" in their representational development (diSessa 2004), it is perhaps surprising that more comparative lists were not included as part of the children's representations. An example of one of the lists created appears in Table 1, with the list comprising words and mathematical symbols.

With respect to compactness, the children continued to leave spaces between bars on their bar graphs, primarily to separate the existing from the desired playground (10 of 14 graphs), and colour was again used extensively on the groups' representations to highlight the information being displayed.

As the children created their representations in both activities, there appeared evidence of their development of conceptual and metarepresentational competence, as noted, especially given that the children had not been instructed formally on the representations they produced.

6.3 Conceptual and Metarepresentational Competence

For activity 2, the apparent links between these competencies were especially evident as the children constructed their circle graphs, given that this form of representation was well beyond their existing curriculum. The children applied their mathematical knowledge in estimating the size of the sectors and their metarepresentational knowledge in building, assessing, and rebuilding their representation, as indicated below.

Each of the focus groups used a ruler and/or estimation in an effort to represent their data. For example, one group argued over how to estimate a sector for each response option, with one child insisting that "You have to find the middle first.

That's the first thing you actually do." He then placed his ruler through the centre of the circle and drew a small sector to represent the two "for exercise" responses to the focus question, "Why do you like the equipment you chose?" He explained, "Two will only be like this (drawing a small sector)... cause it's a very small amount." He then recorded "2" in the sector he had created. When asked how many "pieces of the pie" they needed, the group quickly replied "four, cause there's four of them (response options)." They also commented that the four sectors would not be the same size "because if there's two people, this would have to be a smaller piece to fit two people and a bigger piece to fit nine people in." In estimating the size of the sector to represent the nine responses of "it's challenging," one child claimed that "nine would be half of it" (there was a total of 20 responses to their focus question). After much discussion, the group decided "no, no, no, that nine can't be that big (half of the circle)" taking into account the frequencies of the other response options (six, three, and two). One child subsequently tried to measure the sectors with his fingers to make the nine sectors smaller than the total of the other response options (11), explaining, "Yeah, it actually does have to be a bit smaller."

Another display of the children's conceptual and metarepresentational competence in activity 2 occurred when they commented that their second representation would display the same data. After completing a bar graph, one student group decided to construct a circle graph to display their findings from their focus question, "Why do you like the playground?":

> Peter: You just do the same. It's a pie.
> Belinda: So how many people are there...
> Kim: But we need to do the biggest one (sector) for the most amount of votes.
> Peter: We need to write, we do 12 for the most amount of votes; we take the data that we've got from here (bar graph) and we write what it is.
> Belinda: Challenging (one of the response options) is the biggest.
> Sebastian: So write up "challenging is the biggest" so do it in a pie chart and write 12 percent."...
> Peter: Yeah, we just take the stuff from here and put it there (on the circle graph).
> Sebastian: So do a big thing there and write 12 percent.

What was especially interesting in the children's apparent links between conceptual and metarepresentational competence in activity 2 was their application of percent and percentages, and their efforts in trying to represent 0 %. Of the nine grade 2 student focus groups that progressed to the third year, seven tried to apply their awareness of these ideas in creating their representations. The children had not formally studied this topic but were transferring their learning from their experience with the grade 7 class in the second year (e.g. "Do we have to do percentages like the Year 7s did last year?") and sharing their ideas with those group members who were not involved in the grade 2/7 experience. However, one group who explained that "12 %" means "12 people" subsequently noted that, "We don't know how to use it."

For the five groups who received zero preferences for one of their response options, considerable debate was held on how to display 0 % and 1 %. For example, one group was experimenting with how to represent their data on their circle graph for the response option to the question, "Why do you like the playground?" The response option "fits more people than the oval" received zero votes, while the option "good views" received one preference. After instructing Kim to "Write 12 percent" in one sector of the circle graph (representing 12 responses), Belinda said, "Maybe you could rule a little bit off it... that could be zero percent." She further noted that the response option of good views "only has one percent..." and "has to be really small, like that small." This group further struggled with their display of 0 %, claiming that there was insufficient space to label the option of fits more people than the oval. When their teacher asked, "How can you show 0 %?" Belinda responded that "You should just rub that out... cause that got nothing." But then Kim was puzzled by "How would you do zero?" to which Belinda replied "Rub it out, rub it out."

Another group, however, demonstrated a more advanced understanding of percent, albeit they did not calculate all the percentages correctly. When two group members recommended recording the number of focus question responses (to the question, "Why do you like the Spider Web?") in the circle graph segments they had drawn, Hugh disagreed, saying "Do percent". When the research assistant queried the group on how they intended determining this, they explained that they knew that the circle graph represents 100 % and that of the 20 votes for their focus question, there were two responses that each received five votes, one that received four, and one that scored six. Hugh explained that instead of recording the actual number of focus question responses, percentages should be shown: "That is 100 %, so we needed to do 25 %; that's 25 (%), so that should be 24 (%), and that should be 25 (%) and that one should be 26 (%).

Further evidence of conceptual and metarepresentational competence appeared in the third activity, such as with the group who constructed the "heart monitor" graphs. There was considerable debate on whether such a representation was appropriate for the third activity, with questions raised on using centimetres to display differences in the desired and existing playground features. In Megan's explanation below, it appears her conceptual understanding of using a unit of measurement to represent their data cautioned her against constructing another representation of this nature, even though she was accepting of this in the previous activity. The increase in the number and range of attributes being considered in the third activity might have contributed to her reticence here.

> Megan: You know... how hum like the beep beep one like the doctor (using hand movements to illustrate rises and falls in a line graph), but that wouldn't really work with this, wouldn't it? Like we did lines like 15 people voted for one (response option) so we did a 15 cm line and that was sort of like a doctor's way of doing it. But I don't think that will be good this time.
> James: Well, we could use it.
> Megan: No, it wouldn't really work cause this is not a number. I think a writing one.

Megan thus decided to use "this piece of plain paper representation," to which James disagreed. Megan responded, "That wouldn't show it... you're trying to show the difference between our two playgrounds, how can you show that with centimetres?" Nevertheless, James created a "heart monitor" graph (which didn't show the playground differences), while two other members of the group used his idea to create two such graphs that did show differences in two features, as indicated previously.

Other examples of conceptual and metarepresentational competence in the third activity include instances of children's advanced understanding of scale in constructing their bar graphs. Realising that a scale of 1 square = 1 unit on the 2.5 cm squared grid paper was inadequate for displaying their data, two groups chose to change the scale to 1 square = 10 units. As an example, Tina's group began by discussing the different types of graphs they had constructed during their participation in the study. Tina noted, "I think you can do a line graph, a pie graph, a bar graph… and picture graphs." In deciding to show that their group wanted 105 more on the existing 105 climbing chain squares, they realised that the 2.5 cm squared grid paper was inadequate and suggested "You could divide each square into something" and could also "keep adding paper" (i.e. joining grid sheets end to end). Excerpts from the group conversation below illustrate their conceptual competence in adjusting the scale of their graph and their metarepresentational competence in explaining how they constructed their representation with their new scale.

> Elizabeth: Each square represents 5… is there going to be enough squares? Is there going to be enough squares?

Tina subsequently counted the number of squares in the vertical columns: "Darn it, I'm six short". Elizabeth then counted the squares in fives and arrived at a total of 75 squares, with another child noting, "She's doing each square is five." In completing their representation, Tina explained, "I'm writing a key up here" upon which she wrote, "1 square = 5," but subsequently changed it to "1 square = 5 squares." In a second representation, albeit still a bar graph, the group decided to represent what they would like to be done with the height of the spider web, assuming that "we need to keep the same three colours." Of interest here, is the group's decision to display, "Already" (what we already have), "Total" (of what we want), and "Difference" (between what we have now and what we would like), as indicated in Fig. 3. In reporting back to the class on how many metres high they would like the spider web to be, the group explained that their representation displayed just "one square = one metre" and "We estimated that the spider web was originally 13 metres high and we thought that we would add 12 metres on so that it would equal 25 metres." Unfortunately, the class did not question the practicality of the spider web being 25 metres high. The group further commented that their bar graph representations were "technically the same but different objects."

7 Discussion and Concluding Points

This chapter has addressed data modelling as a means of promoting statistical literacy in the early grades. As a developmental process, data modelling begins with young children's investigations of meaningful phenomena, progressing to identifying various attributes of the phenomena, and subsequently organising, structuring, visualising, and representing data (Lehrer and Lesh 2003). The focus in this chapter

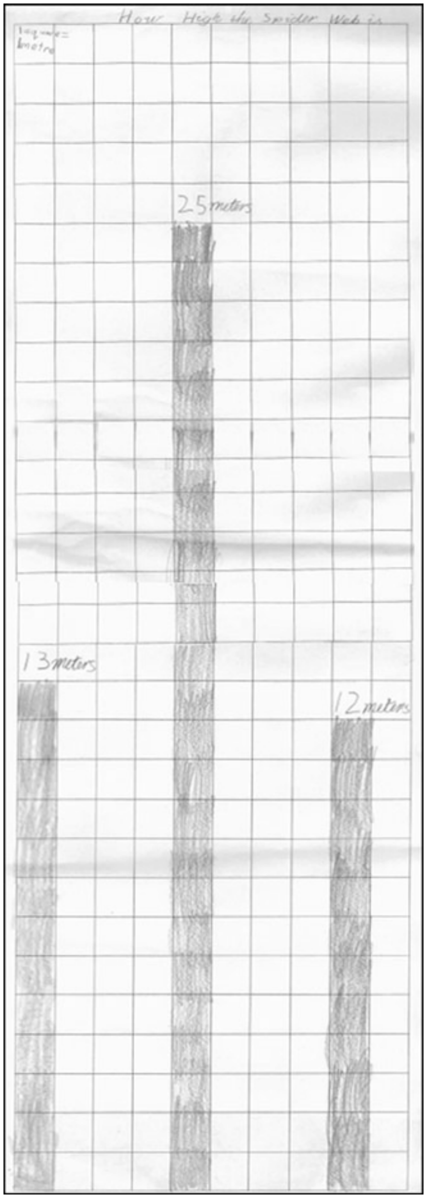


Fig. 3 Example of bar graph displaying existing, total and difference amounts

has been on children's abilities to structure and represent data, and their application of conceptual and of metarepresentational competence in doing so. Children's responses to the last two of the three activities implemented in the final year of a three-year longitudinal project across first through third grade have been addressed here. The activities engaged the children in posing and refining questions; identifying, deciding on, and measuring attributes; developing and conducting a survey;

collecting and recording data; organising, interpreting, analysing and representing data; and developing data-based explanations, arguments, inferences, and predictions.

In working the activities, children generated a variety of multiple representations with bar graphs and circle graphs being the most popular forms. The sophistication displayed in the children's constructions was beyond that demonstrated in their regular mathematics work. This was especially the case for the circle graphs, which do not appear in the school curriculum until later grades. The unusual representations, such as the "heart monitor" graphs, displayed a creative and analogical (English 2004) approach to data representation. Although some might comment that such graphs did not "accurately" portray the data, the children were nevertheless creating their own forms of representation (diSessa 2004) and were not confining themselves to a narrow set of "classroom representations;" they were drawing upon their out-of-school knowledge. Using hand gestures, the children demonstrated how they visualised such a graph would represent their data. The length (in cm) of the lines that formed these graphs corresponded to the quantity being represented, as indicated in Fig. 1.

In activity three, it was interesting to observe the children's realisation that they would have to consider the use of scale in representing the increased quantities with which they were dealing. For example, in the "heart monitor" graph, a scale of 1 cm = 1 m was adopted because the children knew they couldn't use a metre ruler. Likewise, in constructing their bar graph representations, the children realised that taking one square as one unit was inadequate, and hence they decided to change the scale to 1 square = 10 units. It is worth noting that the children had not been formally introduced to the use of scale in graph construction. Neither had they been taught how to construct circle graphs. These examples illustrate the "native capacities" of children, where they create and recreate their own forms of representation (diSessa 2004), applying conceptual and metarepresentational competence in doing so.

Further examples of these competencies were evident in the children's creations of their circle graphs, where they invented a number of ways to determine the placement of their sectors. These ways included using a ruler to find the centre of the circle then estimating the placement of the sectors, for example, using "finger measurements." Some groups applied their understanding (albeit limited) of the notion of percent in determining sector size. The children were aware that the sectors would need to be in proportion to the quantities being represented and engaged in considerable debate on what size the sectors should be. Their discussion here displayed interplay between their conceptual understanding of the comparative quantities being addressed and how the circle graph should be constructed to reflect these quantities. Of particular interest here were children's discussions on how to represent "0 %" and "1 %" on their circle graph; some groups also had difficulty in deciding how a zero amount should be indicated on the vertical axis of their bar graph. There has been very limited, if any, research on these concepts (personal communication J. Whitin, Jan., 2012); clearly more research is needed here.

A number of implications for fostering statistical literacy in the early grades arise from the study reported here. First, children's literature can serve as a rich context for designing and implementing statistical activities. Second, such activities need to be broadened so that children can undertake their own challenging investigations where they select and measure attributes of their choice, and organise, structure, and represent data in ways that they select—such ways might not always conform to "traditional" classroom practice. Third, more activities should be designed where children can create multiple representations, modify and combine representations, and select appropriate representations—all of which are important metarepresentational skills that apply to both mathematics and science (diSessa 2004). Fourth, encouraging creative and innovative ways of working with data in a range of settings enables children to appreciate that data are omnipresent in their world and can be dealt with in a variety of ways.

Finally, incorporating the sharing of model creations with class peers creates opportunities for critical and constructive exchanges of statistical ideas and processes. Furthermore, providing opportunities for cross-grade sharing of statistical activities enable younger children to learn from their older peers, who in turn can become aware of the capabilities and statistical approaches of their younger counterparts.

Acknowledgements This project was supported by a three-year Australian Research Council (ARC) Discovery Grant DP0984178 (2009–2011). Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the author and do not necessarily reflect the views of the ARC. I wish to acknowledge the excellent support provided by the senior research assistant, Jo Macri and the enthusiastic participation of the children and teachers.

References

Clarke, B., Clarke, D., & Cheesman, J. (2006). The mathematical knowledge and understanding young children bring to school. *Mathematics Education Research Journal*, *18*(1), 78–103.

Clements, D. H., Sarama, J., Spitler, M. E., Lange, A. A., & Wolfe, C. B. (2011). Mathematics learned by young children in an intervention based on learning trajectories. *Journal for Research in Mathematics Education*, *42*(2), 167–198.

diSessa, A. A. (2004). Metarepresentation: native competence and targets for instruction. *Cognition and Instruction*, *22*(3), 291–292.

diSessa, A. A., & Sherin, B. L. (2000). Meta-representation: an introduction. *Journal of Mathematical Behavior*, *19*, 385–398.

diSessa, A. A., Hammer, D., Sherin, B., & Kolpakowski, T. (1991). Inventing graphing: metarepresentational expertise in children. *The Journal of Mathematical Behavior*, *10*, 117–160.

English, L. D. (2010). Young children's early modelling with data. *Mathematics Education Research Journal*, *22*(2), 24–47.

English, L. D. (2003). Reconciling theory, research, and practice: a models and modelling perspective. *Educational Studies in Mathematics*, *54*(2&3), 225–248.

English, L. D. (Ed.) (2004). *Mathematical and analogical reasoning of young learners*. Mahwah: Erlbaum.

English, L. D., & Watters, J. (2005). Mathematical modeling in the early school years. *Mathematics Education Research Journal*, *16*(3), 58–79.

Franklin, C. A., & Garfield, J. (2006). The GAISE project: developing statistics education guidelines for grades pre-K-12 and college courses. In G. Burrill & P. Elliott (Eds.), *Thinking and reasoning with data and chance, 68th yearbook* (pp. 345–376). Reston: National Council of Teachers of Mathematics.

Gal, I. (2002). Adults' statistical literacy: meanings, components, responsibilities. *International Statistical Review*, *70*, 1–51.

Ginsburg, H. P., Cannon, J., Eisenband, J. G., & Pappas, S. (2006). Mathematical thinking and learning. In K. McCartney & D. Phillips (Eds.), *Handbook of early child development* (pp. 208–230). Oxford: Blackwell.

Greer, B., Verschaffel, L., & Mukhopadhyay, S. (2007). Modelling for life: mathematics and children's experience. In W. Blum, W. Henne, & M. Niss (Eds.), *Applications and modelling in mathematics education, ICMI Study 14* (pp. 89–98). Dordrecht: Kluwer.

Langrall, C., Mooney, E., Nisbet, S., & Jones, G. (2008). Elementary students' access to powerful mathematical ideas. In L. D. English (Ed.), *Handbook of international research in mathematics education* (2nd ed., pp. 109–135). New York: Routledge.

Lehrer, R., & Lesh, R. (2003). Mathematical learning. In W. Reynolds & G. Miller (Eds.), *Comprehensive handbook of psychology* (Vol. 7, pp. 357–390). New York: Wiley.

Lehrer, R., & Romberg, T. (1996). Exploring children's data modeling. *Cognition and Instruction*, *14*(1), 69–108.

Lehrer, R., & Schauble, L. (2006). Cultivating model-based reasoning in science education. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 371–386). New York: Cambridge University Press.

Lehrer, R., & Schauble, L. (2007). Contrasting emerging conceptions of distribution in contexts of error and natural variation. In M. C. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 149–176). New York: Taylor & Francis.

Lesh, R. A., & Kelly, A. E. (2000). Multi-tiered teaching experiments. In R. A. Lesh & A. Kelly (Eds.), *Handbook of research design in mathematics and science education* (pp. 197–230). Mahwah: Erlbaum.

Lesh, R. A., & Lehrer, R. (2000). Iterative refinement cycles for videotape analyses of conceptual change. In R. A. Lesh & A. E. Kelly (Eds.), *Handbook of research design in mathematics and science education* (pp. 665–708). Mahwah: Erlbaum.

Nobisso, J. (1998). *Hot Cha Cha*. New York: Winslow Press.

Olson, D. R. (1994). *The world on paper*. Cambridge: Cambridge University Press.

Papic, M., Mulligan, J. T., & Mitchelmore, M. C. (2011). Assessing the development of preschoolers' mathematical patterning. *Journal for Research in Mathematics Education*, *42*(3), 237–268.

Perry, B., & Dockett, S. (2008). Young children's access to powerful mathematical ideas. In L. D. English (Ed.), *Handbook of international research in mathematics education* (2nd ed., pp. 75–108). New York: Routledge.

Russell, S. J. (1991). Counting noses and scary things: children construct their ideas about data. In D. Vere-Jones (Ed.), *Proceedings of the third international conference on the teaching of statistics* (pp. 158–164). Dunedin: University of Otago.

Shaughnessy, J. M. (2010). Statistics for all: the flip side of quantitative reasoning. Retrieved 14 August, 2010, from http://www.nctm.org/about/content.aspx?id=26327.

Smiley, N. (2009). *Simon and Catapult Man's perilous playground adventure*. Halifax: Nimbus Publishing Limited.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: ground theory procedures and techniques*. Thousand Oaks: Sage.

van den Heuvel-Panhuizen, M., & Van den Boogaard, S. (2008). Picture books as an impetus for kindergartners' mathematical thinking. *Mathematical Thinking and Learning: An International Journal*, *10*, 341–373.

Watson, J. M. (2006). *Statistical literacy in school: growth and goals*. Mahwah: Erlbaum.

Whitin, D. J., & Whitin, P. E. (2011). *Learning to read numbers: integrating critical literacy and critical numeracy in K-8 classrooms*. New York: Routledge.

Learning Bayesian Statistics in Adulthood

Wolff-Michael Roth

Abstract Probabilities are a pervasive aspect of human life and probabilistic thinking is part of (sociological, psychological) models of rational (social) actors. Yet research documents poor understanding of probability in the general public and suggest that people do not update their estimates about future events when given additional information. This may point to the difficult nature of Bayesian thinking, the framework that would allow rational decision makers to update their estimates about future probabilistic events. Drawing on first-person methods for the study of cognition, I articulate invariants of learning an advanced statistical topic: Bayesian statistics. The first case study focuses on the learning of some fundamentals (such as those that one may find as content of a Wikipedia page); the second case study presents and analyzes a learning episode in the case of quantitative social science research that takes into account prior studies for establishing prior probabilities required for calculating posterior probabilities given the information collected in the study. The analyses show—consistent with pragmatic theories of language—that the essential dimension of learning is what equations, terms, and formulae require to be done rather than their (elusive) "meaning."

> Some statisticians argue that Bayesian statistics cannot be taught effectively in elementary courses, and therefore that it should not be taught. . . . One reason some statisticians believe Bayesian statistics is too difficult for an elementary course is that they do not understand it. (Berry 1997, pp. 241–242)
>
> Similarly, you only understand an expression when you know how to use it, although it may conjure up a picture, or perhaps you draw it. (Wittgenstein 1976, p. 20)

Taking into account information to update the predictions that we make about events in the world appears to be a commonsense issue. Why would people not use additional information to improve their predictions? Updating probabilities based on new information or seeking information to update probabilities requires Bayesian forms of thinking, that is, forms of thinking that increasingly common in everyday life situations (e.g., Lecoutre et al. 2006). Bayesian thinking, as the introductory

W.-M. Roth (✉)
University of Victoria, MacLaurin Building A557, Victoria, BC, V8P 5C2, Canada
e-mail: mroth@uvic.ca

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_24,

quotation suggests, is not easily taught. Others add by stating that "Bayesian reasoning requires a grasp of conditional probability, a concept confusing to beginners" (Moore 1997, p. 254). These quotations pertain to teaching Bayesian statistics at the tertiary level, where students have to take courses; how much more difficult might it be when adults decide to learn (about) this form of thinking when they do not have to? Although the Bayesian approach apparently becomes more widely used, I actually do not know any educational scholar in person—even those who teach the statistics courses at the universities where I taught—who is using or knows how to use Bayesian statistics. But perhaps, as the second introductory quotation intimates, rather than focusing on understanding (mathematical, statistical) expressions and their "meaning," we should focus on their (public) use, which would be acquired in using them together with others.

Probabilities are said to be a pervasive aspect of human life and probabilistic thinking is part of (sociological, psychological) models of rational (social) actors. Yet more than 20 years of research in the learning sciences consistently "documented poor understanding of probability among different populations across different settings" (Liu and Thompson 2007); and there are different kinds of misconceptions even with relatively simple concepts such as correlation (Castro Sotos et al. 2009). Even when statistical methods are available and proven means to ameliorate decision-making, for example, in corporate settings, people "are not aware of the types of rationality underlying the decisions they make, or of the way they could improve their decision-making process by using statistical methods" (Hahn 2011). Workplace research suggests that few non-graduate employees need to draw on statistical reasoning and formal statistical techniques tend not to be at the disposal of employees (Bakker et al. 2008). Strongly held intuitions and common sense heuristics at odds with formal statistics come to be so dominant that they have been used as explanations for the resistance of students to learning probability and statistics (Konold 1995). In addition, psychological studies suggest that people, asked about future probabilistic events, do not update their estimates about the event or, when provided with a choice concerning additional information they would like to have for reassessing their estimates, opt for information that confirms existing hypotheses and does not inform decision-making processes (Nickerson 1998).

In the sciences, too, bias exists and there are many situations where a more appropriate Bayesian approach is not taken (Nickerson 1998). Thus, most research studies in education and the learning sciences that use statistics do not take the results of prior studies into account to inform their own research, treating each study as an independent event even though within the article, authors may suggest that others have or have not shown the same statistical patterns. This may point to the difficult nature of Bayesian thinking, the framework that would allow rational decision makers to update their estimates about future probabilistic events (but just plain folks may not use Bayesian forms of thinking because everyday decision-making is not well modeled by rational actor models Bourdieu 1997). The first introductory quotation points to this fact: even though psychologists tend to assume that reasoning is faultive if we do not reason as Bayesians, most statisticians appear to assume that this form of statistics cannot be taught at the more elementary levels of university education, let alone to just plain everyday people. This quotation also suggests that statisticians themselves may not understand the specifics of the Bayesian approach that makes it so useful.

Drawing on first-person methods for the study of cognition (Roth 2005a, 2012), this chapter is designed to articulate invariants of learning an advanced statistical topic: Bayesian statistics. The first case study focuses on the learning of some fundamentals (such as those that one may find as content of a Wikipedia page); the second case study presents and analyzes a learning episode in the case of quantitative social science research that takes into account prior studies for establishing prior probabilities required for calculating posterior probabilities given the information collected in the study of interest. The analyses show—consistent with pragmatic theories of language (Wittgenstein 1953/1997)—that the essential dimension of learning is what equations, terms, and formulae require to be done rather than their (elusive) "meanings." Because learning to think within a Bayesian framework is like learning an entirely new language, even scientists with statistical background may find it difficult to learn using the Bayesian approach. Even less should we expect just plain folks to use these forms of thinking or fault them for not using these.

1 Method

Although I was trained as a statistician for the social sciences, I did not have the opportunity to learn Bayesian statistics. As a professor, I did not teach Bayesian statistics because (a) it was not part of the curriculum and (b) because I did not know it myself. I might have been one of those statisticians that the first introductory quotation is about—they do not teach it because they do not understand it. I never got to learn Bayesian statistics while working as a statistician and I had moved on to teach qualitative research. Yet I had repeatedly toyed with the idea to know more about Bayesian statistics and even read some articles about the power of using it instead of normal frequentist statistics in social science research. In this chapter, I am not concerned with the particulars of how statistical problems emerged in my work and how I approached them, but in what we can learn about learning (statistics) somewhere along the lifespan more generally. That is, in the case descriptions that I provide below, we have to extract the general and invariant properties that can account for adult learning (of Bayesian statistics) more generally. This chapter, therefore, is not about *my* learning some principles of Bayesian statistics but is about what we can learn about my *learning* Bayesian statistics more generally.

In this chapter, I employ a first-person method. Although first-person methods have come into disrepute in psychology and much of research on mathematical cognition—with the beginning of scientific psychology—widely celebrated (phenomenological) philosophers (e.g., Husserl 1939; Merleau-Ponty 1945) and even neuroscientists (Varela 2001) have drawn on the first-person method to investigate cognition generally and mathematical cognition more specifically. Based on this work, I developed the method of cognitive phenomenology (Roth 2005a, 2005b).

In mathematics education, first-person methods have found their entry, too (e.g., Handa 2003; Roth 2001; Wagner 2011).

The approach used here is very different in its conception than that used by (radical) constructivists—with their fallacious thinking about subjectivity and objectivity. One study on statistical thinking proposes that the researcher is involved in two practices: "constructing an understanding" of what participants do and "constructing a theoretical framework" by reflecting on the activity of coming to understand individuals' understandings (Liu and Thompson 2007). Thus, "when the researcher becomes aware of the concepts and operations upon which (s)he draws to understand individuals' understandings, and can describe them as a coherent system of ideas, (s)he has a theoretical framework" (p. 117). There is a problem, however, which has already been pointed out long ago:

> Suppose we both had the same page of rules in our minds, would this guarantee that we both applied them alike? You may say, "No, he may apply them differently." Whatever goes on in his mind at a particular moment does not guarantee that he will apply the word in a certain way in three minutes' time. (Wittgenstein 1976, p. 23)

Thus, whatever it is that goes on in another person's mind and whatever goes on in the researcher's mind, even if there are the same words and rules of application, does not guarantee that the same use will be made of rule or word. In contrast to what these researchers claim, understanding is not about what is in the mind of others.

> The use of the word "understand" is based on the fact that in an enormous majority of cases when we have applied certain tests, we are able to predict that a man will use the word in question in certain ways. If this were not the case, there would be no point in our using the word "understand" at all. (Wittgenstein 1976, p. 23)

This does not guarantee us a way out of the problem of the general (objective) and specific (subjective), because those concepts and operations may still be particular to the researcher, who would then be led to denote the particular as general. The authors realize this limitation, attributing it to small sample sizes: "This small sample size does not support making claims about the prevalence of this study's central findings to a broader population of teachers" (Liu and Thompson 2007, p. 157).

As the Russian psychologist L. S. Vygotsky realized, even the study of single cases allows us to extract general properties that pertain *to every case*—and he provides an example from his own research where extracted the psychology of art from the study of one case of literature using two others for analogical purposes (Vygotsky 1927/1997). The point of research should be to get at the invariant properties of learning or cognition. To arrive at the general (universal) requires living up to the challenge of *systematically* interrogating "the particular case by constituting it as a 'particular instance of the possible'… in order to extract general or invariant properties that can be uncovered only by such interrogation" (Bourdieu 1992, p. 233). In contrast to the constructivist approach, the particular has to be taken *as* a particular and the point is "to discover, through the application of general questions, the invariant properties that it conceals under the appearance of singularity" (p. 234). One achieves this by completely immersing oneself "in the particularity of the case at hand without drowning in it… to realize the intention of *generalization*… through this particular manner of thinking the particular case which consists of actually thinking it as such" (pp. 233–234). Case-based research too frequently does not lead to generalization and furthermore "inclines us toward a sort of *structural conservatism* leading to the reproduction of scholarly doxa" (p. 248).

2 Learning Statistics in Adulthood

In this section, I present two case studies of learning aspects of Bayesian statistics, with which, as a classically trained statistician, I had not been familiar. At issue in presenting these case studies are not the particulars but what we can learn about the learning of statistics somewhere in adulthood more generally.

2.1 Learning Bayes Formula

2.1.1 Emergence of a Problematic from Pre-Bayesian Concerns

In real-life situations, and in contrast to mathematics lessons, we do not act like mathematicians and are not aiming at acting in this manner. There is no curriculum that suggests specific dates and times when we ought to learn (about) statistics. If there are statistical problems that arise then it is in the course of our everyday concerns for doing some job. Getting this job done sometimes requires us to go beyond what we currently know, the beginning of expansive learning, that is, a form of activity that promises an increase in our control over a segment of our lifeworld.

I got myself into Bayesian statistics without really intending it: during a research project with a colleague who had gathered some data on the changes in different motivational dimensions when students engaged in a specially designed curriculum on green chemistry. I ended up creating tables that showed how many students would exhibit low or high self-efficacy beliefs after treatment when they had exhibited low or high self-efficacy beliefs preceding the treatment. As part of thinking about the traditional χ^2-test, I was wondering about Markov chains and about transition matrices that would allow me to predict the probability to be in the group with high self-efficacy beliefs knowing that preceding the intervention a student had been in the low or high efficacy group.

My first representation involved a table in which I recorded the number of students who had high/low self-efficacy beliefs before and after intervention. But this table did not get me any further. One of the ways in which I represented the results was in terms of an odds ratio, which compared the likelihood to be in the high group after intervention to that before intervention.

$$\frac{\frac{17}{20}}{\frac{6}{20}} = \frac{17}{6} \cong 3$$

What I had calculated in fact was something like an odds ratio that compares the observed frequency to the frequency that would have been observed had there been no change. Such a ratio is of special interest "because an observation gives one little evidence about the probability of the truth of a hypothesis unless the probability of that observation, given that the hypothesis is true, is either substantially larger or substantially smaller than the probability of that observation, given that the hypothesis is true" (Nickerson 1998, p. 177).

As soon as I was thinking odds ratios, somehow the idea emerged to look for probabilities given that there were already probabilities known. I also typed in "Markov" and "chain" into the search engine; the name of Bayes showed up repeatedly. That is, the preceding representations, though given rather than intended, made available resources for new actions that previously were not available. These new possibilities arise from the representations rather than being the result of cogitations. In my research notebook, I enter:

$$
\begin{aligned}
&p(b|a)\\
&\text{initial odds high} \quad \frac{6}{20} = 0.3,\\
&\text{find odds high} \quad \frac{17}{20} = 0.85,\\
&-\text{high} \quad = 0.15.
\end{aligned}
$$

From this, the representation changes to $p(b) = 0.85$ and $p(\text{Б}) = 0.15$, next to which appears the probability

$$
p(\text{high}|\text{high}) = 1, \qquad p(\text{low}|\text{low}) = \frac{3}{14},
$$

$$
p(\text{high}|\text{low}) = \frac{11}{14} = 0.79.
$$

In this situation, a notation system arises that would turn out be consistent with that of Bayesian statistics. What we see here, therefore, is a drift in the activity that self-generated and transformed, producing forms of representations that lent themselves to a subsequent transition to very different concerns. That is, describing this movement as intentional would go too far; but saying that the movement was independent of my previous experience would not recognize the role that the representations used may have had in the emergence of anything like Bayesian ways of thinking. The next entry in my notebook is a transition matrix, which contains the proportion of students with low self-efficacy beliefs who did not change and those who changed to high self-efficacy beliefs. Similarly, I recorded how students changed who had high self-efficacy beliefs at the beginning of the intervention:

	low	high
low	0.21	0.79
high	0	1.00
	0.21	1.79

Following the "transition matrix," which crosses the relative frequencies from exhibiting low/high self-efficacy beliefs before to after, the notes contain another calculation:

$$\begin{array}{ccccc} p(\text{high}) = p(\text{high}|\text{high}) & p\left(\text{high}^b\right) & + & p(\text{high}|\text{low}) & p\left(\text{low}^b\right) = 0.85, \\ 1 & \cdot\ 0.3 & + & 0.79 & \cdot\ 0.7 \end{array}$$

The notation $p(\text{high}|\text{high})$ is a direct translation of the transition matrix into an algebraic form, an implementation of the after|before pattern arising from the issue at hand, the table that explicitly was designed to model the change that occurs as a consequence of a curricular intervention.

At this point, there is no intention yet on my part to learn Bayesian statistics—there is simply a concern for finding a way of representing the particular research results in a particular way. In fact, including the present subsection here was possible only after the fact when it became clear that it constituted the context for my learning of some Bayesian principles. The subsequent concern for learning more about Bayesian statistics *emerged* from the ruminations about an initially unrelated issue, where I wanted to find a way of representing research results in a convincing way. Lifelong learning in mathematics generally and statistics specifically emerges from everyday concerns—as science interests may emerge for people initially not interested simply by reading an (interesting) article about science in an online medium (Roth 2010). But what appeared to have happened is that the framing of the problem and my intense engagement with it primed me for a subsequent engagement to occur. Such a movement is actually intended in the phenomenological epoch: a phase of no attention and accepting of the experiences follows a previous intense engagement with some topic (Depraz et al. 2002). One study in the mathematics education literature cites a pianist and choral director, who suggests this: "The more I play a song, or sing a song, or conduct a song, the more I become attached to it, or it becomes more a part of me. So then, it's as if I develop a relationship with the music—so to speak" (Handa 2011, p. 66). In my situation, too, there was a relation that began to be built that provided the very context for subsequent learning to emerge.

2.1.2 Becoming to Familiar in/Through Use

Later that day and on the following day, I began to browse the Internet using search terms such as "transition AND probability" and "Markov chain." In the process, I also come to a Wikipedia website on "Bayesian inference." This eventually led me to read several introductory chapters in books to Bayesian statistics. Perhaps I am not interested enough at the time, perhaps I experience them as too difficult; the fact is that I am not pursuing the issue to any greater depth. In part, this is mediated by Eq. (1) that is just that, an equation. It has no "meaning" for me and looking at it does not help at the instance.

$$p(H_1|E) = \frac{p(E|H_1)p(H_1)}{p(E|H_1)p(H_1) + p(E|H_2)p(H_2)}. \qquad (1)$$

At the time I leave it to that.

As my research notebook evidences, two days after the initial work on the change matrices, the concern for understanding changing probabilities when information becomes available returns to my mind: in an unpremeditated fashion. I am walking on the beach with my wife. All of a sudden I find myself thinking about a game of two dice. I literally have the image of two dice covered by a hand or piece of cloth. I wonder about the probability of throwing a sum of 7 and how the probability changes when I only get to see one of the two. I find that whatever the first dice shows, the possibility to have the two numbers add to seven will be $1/6$. I realize that the issue is perhaps framed inappropriately and that I need to ask the question about the probability to have the conduction 2 dice: $\sum > 7$. I realize at some point that I have to look for $\sum > 7$ rather than for $\sum \geq 7$. I count out in my head. When the first dice has a 1, then the probability is 0. But in the case of a 2, the probability to have $\sum > 7$ is $1/6$—i.e., $p(\sum > 7) = 1/6$—as only throwing the 6 gives $\sum > 7$. I continue: If the first dice shows a 3, then having a 5 or 6 on the second one would yield $\sum > 7$; if the first dice shows a 4, then having a 4, 5, or 6 on the second would yield the sum. Immediately testing a 6 on the first yields the odds of $5/6$. It then come to me that the total probability of having a sum greater than 7 before looking at all is $1 + 2 + 3 + 4 + 5$, which I add up in my mind's eye to be 15. So $p(\sum > 7) = 15/36$ and, equivalently, $p(\sum \leq 7) = 21/36$.

I realize that the prior probability to have $\sum > 7$ is $p(\sum > 7) = 15/36$, which decreases to $p = 0$ when the event $E = 1$ on the first dice. I now have a sense for the claim that information changes the probability. I am thinking about getting home and playing this out on paper.

At home, taking a letter-size sheet of paper sideways, I begin to note what I had previously envisioned while walking on the beach:

$$2 \text{ dice}, \; p, \; \text{sum} > 7 \; (2,6)\; (3,5)\; (3,6)\; (4,4)\; (4,5)\; (4,6)\; (5,3)\; (5,4)\; (5,5) \\ (5,6)\; (6,2).$$

I add "$H_1 =$" and below it "H_0 sum ≤ 7." I count out the possibilities from the list of number pairs, which turns out to be 15, as I earlier calculated. This counting out is actually a way of confirming or checking that what had been calculated earlier mentally turns out to be the case:

$$p(H_1) = \frac{15}{36}, \qquad p(H_0) = \frac{21}{36}. \tag{2}$$

Looking at the table of values, I then begin to write

$$p(H_1|E) = \frac{p(E|H_1)\frac{15}{36}}{-\frac{15}{36} + -\frac{21}{36}} \tag{3}$$

where the slots for the $p(E|H_1)$ and $p(E|H_2)$ in Eq. (1) are initially left open. I wonder what these values might be. My gaze sweeps over the sequence of number pairs and all of a sudden I realize that there are only two 2s. This means that under hypothesis 1, there are only 2 out of 15 cases where there is a 2. So the probability to have a 2 given H_1 would be 2 out of 15. I complete Eq. (3), which yields

$$p(H_1|E) = \frac{\frac{2}{15}\frac{15}{36}}{\frac{2}{15}\frac{15}{36} + \frac{10}{21}\frac{21}{36}} = \frac{\frac{2}{36}}{\frac{12}{26}} = \frac{1}{6}. \tag{4}$$

This is precisely what I had figured out before: If the first dice shows a 2, then there is a $1/6$ probability that $\sum > 7$ (i.e., 8). I check the formula if the first dice shows a 1, and immediately realize that in this case $p(E|H_1) = 0$, because there is no 1 in my list of pairs. The probability $p(H_1|E) = 0$, just as I had it envisioned.

This all seems to fit together. And then: I realized that I was actually checking in this way that I got the proper values for $p(E|H_1)$ and $p(E|H_2)$. In fact, there are three empty slots in the equation, two of which are to be taken by the same probability. These slots invite certain values. Thus, thinking about what the values possibly are happens on the part of the representation rather than deriving in some willy-nilly fashion from my mind. The consideration still concerns what these values might be, and again, the particular representation chosen of all the pairs of dice values that would yield $\sum > 7$ invites the particular choice made. That is, the list of pairs provided the opportunity to discover the two occurrences of "2" and the absence of a "1," both facts of which led me to the convergent results from the formal and informal approaches.

At this point, a sense of satisfaction emerges. There is a feel of understanding something that has never made sense to me before. But it is not understanding in a "cognitive sense"; rather, it is an understanding that pertains to knowing what to do and that doing something in different ways yields the same results. I am thinking about what was involved in this learning process and about the psychological research that sneers at everyday people who do not use a Bayesian approach to update probability estimates given new information. I realize that it has taken the concrete playing through with an example that I not only could relate to but that has emerged into my mind and that I have been playing around with to get to the point of developing a deep understanding of what before that had been just a formula. In fact, now that I knew *what to do*, I also had a sense of understanding; I am immediately thinking about Wittgenstein and his adage that words do not have meaning, that meaning is irrelevant and what matters is word use.

2.1.3 Analysis

Even though and perhaps precisely because the example is trivial, we can draw several important lessons about learning in everyday situations. In this situation, the particular thought that got me to work and eventually led to the increasing familiarization with the Bayesian formula came to me in an unpremeditated fashion. The thought of the game with the covered dice appeared in my mind's eye while I am enjoying a walk in a beautiful natural setting. But I am a willing host, accepting the thought even though it was so foreign to the setting. In fact, phenomenologically oriented first-person research methods suggest such times to be very useful in arriving at new understandings—it is an integral part of the phenomenological epoch. Epoch involves three phases (Depraz et al. 2002): (a) an initial phase during which experiences are systematically produced all the while suspending one's beliefs about them, (b) a conversion phase during which attention is changed from the content of experience to the process of experience, and (c) a phase of accepting experience (no-attention). The game with the covered dice not only allowed me to build this extended experience but also followed other experiences.

Precisely because I do not know what I will know, I cannot aim at learning what I will eventually know or what someone who has given me a task might suggest I am learning by engaging with it. At the time that the game appears in my mind's eye, I do not even think about Bayes formula or about applying it. I am trying to understand what happens to the probabilities once one of the two dice is revealed. There is nothing that I *can* know that I will ultimately come to a better sense for Bayes formula. That is, both the topic (the game), the problem that arises from it for me, and the ultimate solution *are given to me*—in other words, they are not the consequences of an intentional learning act. In fact, at the time I cannot intend what I will be learning precisely because I have to learn it (how to use or understand Bayes formula) first before it can become the object of my intention. There are in fact analyses of the problem that show that any creator

> is typically unable to make clear exactly what it is that he wants to do before developing the language in which he succeeds doing it. His new vocabulary makes possible, for the first time, a formulation of its own purpose. It is a tool for doing something which could not have been envisaged prior to the development of a particular set of descriptions, those which it itself helps to provide. (Rorty 1989, p. 13)

Precisely because I do not know what I will know, I cannot know whether what I am doing is right. There are certain ways of looking for consistency, such as when I calculate the posterior probability of having $\sum > 7$ once I know that one of the two dice shows a 2. The situation is easy enough, because only a 6 on the second dice will give me $\sum > 7$. The literature on metacognition *cannot be right*, for in the process, I have no idea where all of this is going to end up and whether anything I am doing is correct for getting me to any specific point. All I can expect is this: I do something to see what it gives and then, from the results of my doings, work backwards to reflect on the intentions for doing what I have done and the nature and quality of the steps I have taken to get me where I am.

When I am done, I realize that what had been difficult to envision when I first encountered Bayes formula was what to do when it states $p(E|H_1)$ and $p(E|H_2)$. In my notation, these were precisely the missing values. It was because I had written out all the possibilities of the two dice when $\sum > 7$ that it came to me that there were only 2 possibilities. I did not intentionally write out the pairs of dice values that would give me $\sum > 7$. I realize, above all, that my thinking about the two required probabilities is likely correct, as the result I obtained is correct given that I could get at it given the particularity of the simulation. In this instance, there are two routes that get me to the same result: the one in which I enter quantities in a formula, whereby I am uncertain about a particular expression; and the other one by practical reasoning from understanding the system at hand (once I know that one dice has a 2, I also know that I need a 6 to get a sum greater than 7, for which I know the probability to be $1/6$).

The particular case of the dice fortuitously contains important features: the sequential change of probabilities, which, already without a pencil and paper or a computer available, I realize to change from $15/26$ to $1/6$ once it is revealed that the first dice shows a 2.

2.2 The JZS Bayes Factor in Two Contexts

2.2.1 Episode

Two months after the episode with the dice, I was working again with the colleague on another paper. She had wanted to do an ANOVA study comparing environmental attitudes among three ethnic groups in Malaysia. I wanted to add another analysis testing the degree to which the hypothesis of gender effects was affected by the results from the present study. That is, rather than reporting the standard frequentist probability $p(\text{data}|H_0)$ (the probability that the data have arisen by chance), I wanted to test $p(H_1|\text{data})$—or conversely $p(H_0|\text{data}) = 1 - p(H_1|\text{data})$. I knew that there was a tool available online allowing me to find the JZS Bayes factor

$$BF = \frac{p(\text{data}|H_0)}{p(\text{data}|H_1)} \tag{5}$$

where $BF > 1$ means that the evidence supports H_0 whereas $BF < 1$ is interpreted as evidence for H_1. (Readers will immediately notice that the JZS Bayes factor is the same kind of fraction in my first example, where I wanted to find out the ratio of students with belief change compared to no-change.) I wanted to go through the literature one study at a time to see what authors had reported as t values (plus N_{MALE} and N_{FEMALE}), which I would then use to update the priors sequentially until I got to our study. Going through the literature I realized that some of the studies reported correlations rather than t-tests. But then I found that there was also a tool to calculate the Bayes factor knowing the number of participants N and the correlation r.

The thought came into my mind that I should check whether the two tools would give the same result if I calculated both r and t statistics using the same "data." I created a simple example in a spreadsheet with the independent variable gender taking on values of 0 and 1 so that I would get the following data pairs:

$$\text{Data} = \{(0, 1.1)(0, 0.9)(0, 1.2)(0, 1.3)(0, 0.8)(1, 1.7)(1, 2)(1, 1.8)(1, 1.9)\}.$$

The spreadsheet functions yield $r = 0.9308086$ for the correlation and $p = 9.21854\text{E}{-}05$ for the t-test. Using the TINV function, I find $t = 7.202940576$. This still does not make the two statistics comparable. My search for an equivalent function to find a p-value associated with the correlation yields nothing. But online I find an *r-to-p* calculator, which yields $t(8) = 7.2$, $p < 0.0001$.

Now that I am certain that the two forms of analysis (correlation, t-test) of my data are equivalent, I enter the required values into the parts of the same website corresponding to the calculation of JZS Bayes Factor for t-test and correlation. The results are $JZS_{\text{T-TEST}} = 0.0041709$ and $JZS_{\text{R}} = 0.0025704$. I cannot figure out what is going on here and fire off an e-mail message to J. Rouder, the owner of the website and tools.

> I was playing around with your calculator. I know that people often code gender differences as a correlation using the dummy coding 0,1. I used a little dataset attached. When I do a t-test and a correlation, and use the JCS calculations for 2-sample t, on the one hand, and an R^2, on the other hand (1 variable, square of correlation), I arrive at different JCS factors. Is there something wrong in my thinking that leads me to get different numbers?
> (e-mail, W.-M. Roth, August 30, 2011)

Within a couple of days, I received a response.

> Thanks for your query. Your intuition that they should be the same is correct. And you are correct that they are not. The t-test and regression case are implemented with slightly different priors, with a sociological/historical explanation. In the t-test case, which we did first, we parameterized the distance between the two levels themselves and put a Cauchy prior on this difference. In the regression case, we parameterized difference between each level mean and the grand mean, and put a Cauchy prior on this difference. This difference is $1/2$ the size as the corresponding difference in the t-test. I actually like the first one better, but it is harder to do with more than two levels, and is not what Liang et al. did. I was trying to stay close to Liang et al. in the regression calculator.
> So, let's take your case:
> 0 Condition: 1.1.9 1.2 1.3.8
> 1 Condition: 1.7 2 2.1 1.8 1.9
> Summary statistics:
> mean 0: 1.06
> mean 1: 1.90
> t-value $= -7.2$
> JZS BF from two-sample t-test is about 240:1 in favor of the alternative
> $R^2 = 0.8664$
> JZS BF is about 389:1 in favor of the null
> Now, let's adjust the scale (r) in the one-sample test so that the priors are the same.
> Set $r = 2$. Now, you see that the t-statistics and R^2 statistic yield the same JZS Bayes factor.
> I wish we had been more consistent in our priors across papers. If I had to do it over again, I would use $r = 2$ as the default in the t-test applications, but it is water under the bridge now. I might add a web page explaining this.
> (J. Rouder, September 1, 2011, personal communication)

2.2.2 Analysis

The definition of the JCS Bayes factor parallels what I had done earlier when I compared the probabilities of exhibiting high versus low self-efficacy beliefs described in the preceding subsection. This is not to say that I had intended this, because beforehand, I could not have known about the similarities. The fact that I became aware of the similarity does not mean that this would be so for every person. Moreover, I cannot say why I would have wanted to check whether calculating the JZS Bayes factor for the same data analyzed using Pearson r and t-test. I could have just trusted that the results of the two online tools would be equivalent, especially because the same researchers had built both. In part, there may have been something like a concern of not mixing apples and oranges—there was just a vague unease about using the JZS Bayes factor using two different tools and combining these results in my own repeated recalculation of the priors.

In this situation, we see that when the tools do not provide convergent results, then a problem is identified. This is the converse of the first part of the case study, where the convergence in the results provides some assurance that the conceptual underpinning is correct about the similarity between t-test and Pearson r when the independent variable is dichotomously (bi-valued). The convergence criterion was also seen in the first case study, where the two approaches to calculating the updated probability for looking at the second dice yielded the same result. In each situation, it is the equivalence of the results that allows us to be more confident about the process that produced the results. An analogy from everyday life may be the cookbook that provides photographs of the finished product so that persons following the recipe can compare their results with a reference result. In my situation, I produced both results, and the convergence provides positive feedback.

In the absence of somebody who can provide an evaluation of what we are doing—something like a cultural "enforcer," such as a teacher or other feedback mechanism—self-learners (autodidacts) have to have some mechanism that the results of their actions are likely to be consistent with the cultural forms. When the issue is "meaning" or "understanding," then such mechanisms are not inherently available, because everything happens in the mind, at the ideal level. As Wittgenstein suggests in a quotation provided in the introductory section, we can never know whether two individuals use the same word or rule in the same way; and this is the case for the same person using the word or rule only a few minutes later. But external critique of the mind is possible only when it externalizes itself, that is, makes some product of itself the object (Lat. *ob-*, over and against, *jicere*, to throw). This is beautifully illustrated in Watson's account of the discovery of the double-helix structure of DNA, where the researchers played around with cardboard and metal models of the molecules they knew to be constituent parts. When on the top of the table, they were actually seeing the outcomes of their actions, which, in particular combinations of the bases turned out to be identical pairs (the rungs of the DNA ladder). Thus, actions where there are concrete outcomes can be evaluated in terms of these outcomes—for example, when computing devices are used to do mathematics. (During one research project, I gave fifth-graders my laptop with a mathematical modeling program, which allowed them to learn how to get equivalent fractions because the computer provided them with feedback about the correctness of fractions they compared.)

We notice that in this as in the preceding situation, a concrete case plays a central role. This case is simple enough, as is the one with the two dice, and would therefore allow, if required, hand calculation. But this simple example allows for a quick and rapid test of alternative tools, which, in this situation, yield the same results in one instance (same t and p values) and different results (JZS Bayes factors). Unlike in the second case, there is a person (J. Rouder) who provides feedback and therefore a resolution to the problematic situation at hand. These simple situations also are correlated with the feeling of the “sense of the Bayesian approach,” much in the way mathematics educators know about and research number sense: a “feel for number” (Wagner and Davis 2010). To have a “feel for” means to have an (intuitive) understanding or practical sense of how something works. In other words, it means knowing one’s way around a particular part of the world. As the present case studies show, the concrete examples allow a feel for the systems under investigation to emerge, a way of knowing one’s way around this particular aspect of the world. There is no longer a difference between knowing the language of Bayesian statistics and knowing one’s way around problems that afford the use of Bayesian statistics more generally.

3 Learning and Teaching Statistics in Adulthood

The advantage of the present approach lies in the fact that it can be used for developing strategies for teaching and learning (Bayesian) statistics at any place along the trajectory of life. But if we were to focus on “meaning” and “understanding” that reside somewhere in the mind of people and are only denoted by the words and signs they use, there would really be little that we can do, as it has to be up to the people themselves to construct whatever is in their mind. Tool use, on the other hand, occurring in the marketplace of public discourse and interaction, can be taught through emulation, trial, and feedback.

3.1 From Knowing/Knowledge to Doing

The issue I am grappling with here is similar to the one that a language philosopher has identified for the learning of any language. It

> depends on the assumption that communication by speech requires that speaker and interpreter have learned or somehow acquired a common method or theory of interpretation—as being able to operate on the basis of shared conventions, rules, or regularities. The problem arose when we realized that no method or theory fits that bill. (Davidson 1986, p. 446)

The same can be said about learning what something like the Bayesian formula “means,” for there is no theory of interpretation or common method that would tell us how such meaning could be constructed “on the basis of shared conventions, rules, or regularities,” for the very interpretive method by means of which students are to make sense have to arise from culture. In the constructivist approach described in the section on method, no such process could get started, for the very means of interpretation that students require are those that exist in culture—because otherwise nothing that they construct would make sense to somebody else.

In my case account, we see how feel for the systems at hand emerges in the same way as competence arises for novice drivers from driving around a parking lot. In

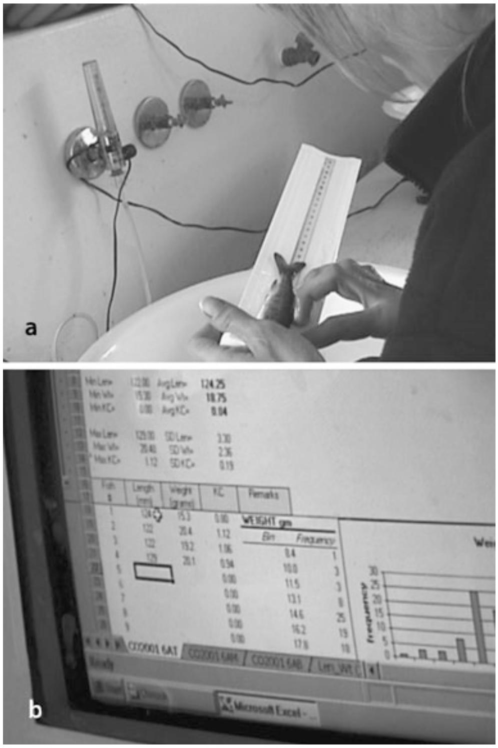


Fig. 1 (**a**) A fish culturist handles fish, weighing and measuring length. (**b**) She enters the measurements and receives immediate feedback as to the distributions of weight and length, means, and calculated factors such as the condition coefficient K_C

the context of a very simple setting, they gain basic competencies while having some assurance that errors will not be fatal. They develop a basic sense for how the vehicle behaves in response to their actions and possibly what the limits are to the actions before dangerous situations will evolve, such as sliding.

Between 2000 and 2004, I had spent two to three days per month in a fish hatchery both observing fish culturists at work and participating in the work myself. During this research, I had an opportunity to study the statistical ways of one fish culturist, who, despite having dropped out of college and despite little mathematical training, looked at the weight and length distributions from a sample of 100 fish and told me what the fish must have looked and felt like. That is, despite looking only at some distributions, she could say that there was a bimodal sample characterized by short fat and long skinny fish. This intimate feel for the relation between statistical representations and their real world equivalents was arising from the close relation between the practical handling of the fish, such as taking length measurements (Fig. 1(a)), the entering of the measurement in the database, and seeing the result that this measurement has on the mean, the distribution, and on such parameters as condition coefficient ($K_C = \text{weight} \cdot 10^5/\text{length}^3$) (Fig. 1(b)). In this situation, too, there is a direct correspondence between an action (measurement, entering it in the computer) and the result (changes in means, standard deviation, condition coefficient [K_C], mean K_C, and the weight and length distributions).

In the present instance, the formalisms used are particular to the case (i.e., to me), given that most students required taking applied statistics courses display "strong reluctance… to any kind of formalism (even the simplest such as the summation sign $\sum$)" (Vidal-Madjar 1978, p. 383). On the other hand, when students do have prior relevant experiences, such as those enrolled in a Masters degree business program with prior field experience, then they actively draw on these experiences as a way of making sense of new situations rather than on formal statistical knowledge: the pragmatic knowledge deriving from their work as salespersons tend to prevail even when the tasks have school-like nature (Hahn 2011). In other school contexts, students develop a feel for statistics and statistical representations—such as frequency distributions (Bakker and Hoffmann 2005)—when they have had extensive experiences with a phenomenon.

The case studies also show a considerable extent of experimentation with practical situations and representations. Because such experimentations yield concrete results, feedback becomes available. Moreover, there is a particular kind of rationality embedded in the use of representations, which make possible actions and transformations as well as place constraints on the actions (Hoffmann 2004; McGinn 2010). Experimentation therefore allows relations to emerge that are heretofore undisclosed to the person, who continually comes up against the hard facts of the concrete results. From a didactical perspective, however, there remains a major problem: "*how to generate* new representational systems, or new elements of those systems, when the limits of our representational means are known" (Hoffmann 2004, p. 302). The author suggests in this context that the idea of creative abduction can lead us further. This is a process where the agent generates a possible solution to the generalization of a problem and then figures out its implications. When the implied results coincide with the actual case, a rule has been found that may constitute the appropriate generalization. In the above-described case that led to the appropriate use of the probabilities $p(E|H_i)$, we have had just such an instance. As the result led to the same probability $p = 1/6$ as the one that resulted from the intuitive analysis of the situation, the actions of calculating $p(E|H_1)$ and $p(E|H_2)$ were accepted as (likely) true.

3.2 Expansive Learning

The case studies show how statistical learning emerged without express intention. This situation differs from that which we would find if persons decided to take a course in statistics or when a course of study has statistics as a basic requirement—as one often finds in graduate programs, where many students enroll only reluctantly. In my two cases, there is not an interest in the object in itself but the engagement in Bayesian statistics emerges in the course of working on a paper publication and the search for a form of representing results that would make the study more convincing to the reviewers and readers. That is, there was a sense that the engagement in Bayesian statistics would expand the action possibilities with respect to the task at hand and, therefore, a greater control over the object/motive—paper accepted for publication.

In this formulation, it is evident that engagement promises or has as possible outcome an expansion of action possibilities. Such learning is denoted by the term *expansive learning* (Holzkamp 1993); it is a concept particularly useful in adult and lifelong learning (Langemeyer 2006). Because of the positive affective values that come with greater control over one's life conditions and with an expanded room to maneuver (range of action possibilities), the concept of expansive learning makes superfluous the concept of motivation (which critical psychologists conceive of as a bourgeois tool for making subordinates do what they do not inherently do on their own) (Roth 2007). One does not have to argue that I (the author) am particularly internally self-motivated; all we need to see is that there is control over the issue at hand and increasing room to maneuver that are the real outcomes that have affective value. Thus, we do not require self-motivation as a concept to explain the sustained engagement in learning (about) Bayesian statistics even though it—and the Bayesian formula in particular—made little sense.

An expansion of action possibilities arose precisely at the point when the formula as a whole and its individual parts found their place in my action repertoire. Thus, rather than thinking about "meaning" or "understanding" that mysteriously attached themselves to Bayesian statistical terms and formulas, familiarity and appropriate use appear to be better ways of conceptualizing the events. Moreover, the expansion in agency as new forms of actions become available and part of the agent's repertoire fit together with the idea of expansive learning that inherently includes the equivalents of "self-motivation" and "internal motivation."

Some educators suggest the replacement of informal forms of statistical knowledge with a more normative one (e.g., Konold 1995). This way of framing the issue, however, is problematic precisely because the informal intuitive ones are grounded in the everyday of knowing one's way around the world. It does not allow us to understand how someone without appropriate conceptions nevertheless comes to learn these. How does one reconstruct one's own misconceptions when the cognitive tools, materials, and ground are consistent with misconceptions? Although some scholars propose that prior experiences possibly are a problematic starting point for teaching statistical concepts (Noss et al. 1999), I suggest that an appropriate way of formulating the issue is dialectical and consistent with the history of mathematics. Thus, the Greek developed formal geometry *on the basis of* their intuitive understandings about the three-dimensional world; their non-formal understandings constituted both the ground and the tools for the emergence of formal geometry (Husserl 1939). In the present context, therefore, we need theory and appropriate curricula that understand intuitive (naïve) understandings of statistics as the ground, material, and tool from/with which more formal statistical knowledge is to evolve. This is consistent with the observation that the use of statistics requires both statistical and disciplinary knowledge (McGinn 2010); and it is consistent with the described study in the fish hatchery, where statistical feel arose from the intimate familiarity with the environment and the effect particular measures have had on the representations. The appropriate theoretical learning concepts are dialectical, because they include, in the category of learning, a contradiction that combines informal/non-formal and formal knowledge into the same unit of transition.

The idea of expansive learning is consistent with the pragmatic approach that replaces the epistemological focus from "meaning" and "understanding" to acting and using representations (tools). It is also consistent with the idea whereby our current familiarity with the world constitutes the basis, tool, and material for developing new ways of navigating this world, which arises from our greater control over it and the expansion of action possibilities. Rather than replacing what we know from participating in the everyday world, I suggest we value our familiarity with this world, our control over this world, and our current action possibilities within it because these are the starting point and ground on which to stand. Any attempt in replacing it would pull the rug from under our feet.

3.3 Dialectic of Agency and Passivity

The preceding section about increasing agency and action potential is only part of the story of learning. We notice in this study how much passivity—expressed in the passive formulations of how ideas came to me, how something or some text led me to something else, and so forth—comes into play together with agency. Ideas come to the thinker rather than being actively produced; but the thinker may accept or reject these ideas. The image of the dice game appeared all of a sudden and precisely at a moment when I was not thinking about it; I developed a feel for the Bayesian formula—although this was not my intention—because of my feel for this game of dice; specific representations that offered themselves up subsequently suggested some rather than other actions and therefore shaped the ultimate learning path; the notation for Eq. (1) emerged from the particular tabular form in which I had represented (unintentionally) the results of our experimental study. I was "playing around" with the representation and was pursuing their affordances; and in so doing, the results were giving themselves independent and despite any intentions that I might have had. Although the specific objects given in this learning process are particular to the cases at hand, the conditions are principled ones. Thus, it is impossible for the learner to intend the learning object precisely because it is the outcome of the process and therefore cannot be orienting its intention; even if there is a place for intention in the processes described, the intentions themselves are unintended, offer themselves to the host who is only in the position of the receiver, who may accept or reject the gift.

The preceding considerations, therefore, lead us to a more differentiated conception of learning (statistics) anywhere along the lifespan. In this conception, learning is in excess of agency because of the excess of intuition over intention that arises from the inherent relation between knowledge and the object of (learning) activity. Learners therefore have to act even though they cannot know what will be the result—as seen in the situation of the first calculation of the posteriors for the dice game. The chosen action—taking as $p(E|H_1)$ the number of 2s in the list of dice point pairs—leads to a particular result. And in the way a painter steps back to see what a brush stroke gives, I was able to see that the result of my actions that I did not know whether they were right or wrong, good or bad. In this case, the result coincided with the one I had obtained following a different route, so that the current action yielded (passive formulation!) a result that could be positively assessed.

4 Coda

Research articles about understanding statistics in the general public tend to decry the "poor understanding of probability among different populations across different settings" (Liu and Thompson 2007, p. 113). As a result of the present study, I suggest that practical use (usability), familiarity, expansive learning (and the concomitant concepts of control and action potential), and the agency | passivity dialectic as alternative set of concepts for thinking about the state and learning of statistics in the general public. There is no reason to believe that the anticipation of expanded control over life conditions and action possibilities in the world would not lead members of the general public to make greater use of statistics. There is little in the everyday world that would suggest any advantage to come from knowing more statistics. Even as a trained statistician, I have very little use for statistical thinking in my daily life (other than for critiquing many of the medical studies that make it into the news). Envisioning greater control and action possibilities, however, requires familiarity with the world. In contrast to other frameworks for learning, mine suggests grounding any statistical learning in everyday phenomena where people can bring and apply their intuitions about the system under investigation—such as the game of dice in my own case. But it should be evident that not every system will do the trick for every person. The person has to be able to build the mentioned relation, so that there may be as many different systems as there are students in a class taking statistics.

In the end, I find myself in the same/similar position as the author of an introductory text on Bayesian reasoning (also a physicist). After noting that the first examples that he did "were so trivial that anyone with a bit of imagination could have solved them just by building contingency tables, without need to resort to probability theory and that 'strange theorem'" (D'Agostini 2003, pp. viii–ix), the author goes on to note that he "was a bit disappointed to learn that [his] wheel had already been invented centuries earlier" (p. ix). But in the end, the self-learning process is recognized as something crucial because consistent with his intuitive ideas:

> I consider this initial self-learning process to have been very important because, instead of being "indoctrinated" by a teacher, as had happened with my frequentistic training, I was instinctively selecting what was more in tune with my intuitive ideas. (p. ix)

The point, however, is not that we merely have reinvented the wheel. This reinvention is in fact the very condition of the historicity and objectivity of mathematics (Husserl 1939).

References

Bakker, A., & Hoffmann, M. H. G. (2005). Diagrammatic reasoning as the basis for developing concepts: a semiotic analysis of students' learning about statistical distribution. *Educational Studies in Mathematics*, *60*, 333–358.

Bakker, A., Kent, P., Derry, J., Noss, R., & Hoyles, C. (2008). Statistical inference at work: statistical process control as an example. *Statistics Education Research Journal*, *7*(2), 130–145.

Berry, D. A. (1997). Teaching elementary Bayesian statistics with real applications in science. *American Statistician*, *51*, 241–246.

Bourdieu, P. (1992). The practice of reflexive sociology (the Paris workshop). In P. Bourdieu & L. J. D. Wacquant (Eds.), *An invitation to reflexive sociology* (pp. 216–260). Chicago: University of Chicago Press.

Bourdieu, P. (1997). *Méditations pascaliennes [Pascalian meditations]*. Paris: Éditions du Seuil.

Castro Sotos, A. E., Vanhoof, S., van den Noortgate, W., & Onghena, P. (2009). The transitivity misconception of Pearson's correlation coefficient. *Statistics Education Research Journal*, *8*(2), 33–52.

D'Agostini, G. (2003). *Bayesian reasoning in data analysis: a critical introduction*. Singapore: World Scientific.

Davidson, D. (1986). A nice derangement of epitaphs. In E. Lepore (Ed.), *Truth and interpretation* (pp. 433–446). Oxford: Blackwell Sci.

Depraz, N., Varela, F., & Vermersch, P. (2002). *On becoming aware: steps to a phenomenological pragmatics*. Amsterdam: Benjamins.

Hahn, C. (2011). Linking academic knowledge and professional experience in using statistics: a design experiment for business school students. *Educational Studies in Mathematics*. doi:10.1007/s10649-011-9363-9.

Handa, Y. (2003). A phenomenological exploration of mathematical engagement: approaching an old metaphor anew. *For the Learning of Mathematics*, *23*(1), 22–28.

Handa, Y. (2011). *What does understanding mathematics mean for teachers?* New York: Routledge.

Hoffmann, M. H. G. (2004). How to get it. Diagrammatic reasoning as a tool of knowledge development and its pragmatic dimension. *Foundations of Science*, *9*, 285–305.

Holzkamp, K. (1993). *Lernen: Subjektwissenschaftliche Grundlegung [Learning: subject-scientific foundations]*. Frankfurt/M: Campus.

Husserl, E. (1939). Die Frage nach dem Ursprung der Geometrie als intentional-historisches Problem. *Revue Internationale de Philosophie*, *1*, 203–225.

Konold, C. (1995). Issues in assessing conceptual understanding in probability and statistics. *Journal of Statistics Education*, *3*(1). Accessed December 22, 2011 at http://www.amstat.org/publications/jse/v3n1/konold.html.

Langemeyer, I. (2006). Contradictions in expansive learning: towards a critical analysis of self-dependent forms of learning in relation to contemporary socio-technical change. *Forum Qualitative Sozialforschung (Forum Qualitative Social Research)*, *7*(1). http://nbn-resolving.de/urn:nbn:de:0114-fqs0601127.

Lecoutre, M.-P., Rovira, K., Lecoutre, B., & Poitevineau, J. (2006). People's intuitions about randomness and probability: an empirical study. *Statistics Education Research Journal*, *5*(1), 20–35.

Liu, Y., & Thompson, P. (2007). Teachers' understandings of probability. *Cognition and Instruction*, *25*, 113–160.

McGinn, M. K. (2010). Learning to use statistics in research: a case study of learning in a university-based statistical consulting centre. *Statistics Education Research Journal*, *9*(2), 35–49.

Merleau-Ponty, M. (1945). *Phénoménologie de la perception*. Paris: Gallimard.

Moore, D. S. (1997). Bayes for beginners? Some reasons to hesitate. *American Statistician*, *51*, 254–261.

Nickerson, R. S. (1998). Confirmation bias: a ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*, 175–220.

Noss, R., Pozzi, S., & Hoyles, C. (1999). Touching epistemologies: meanings of average and variation in nursing practice. *Educational Studies in Mathematics*, *40*, 25–51.

Rorty, R. (1989). *Contingency, irony, and solidarity*. Cambridge: Cambridge University Press.

Roth, W.-M. (2001). Phenomenology and mathematical experience. *Linguistics & Education*, *12*, 239–252.

Roth, W.-M. (Ed.) (2005a). *Auto/biography and auto/ethnography: praxis of research method*. Rotterdam: Sense Publishers.

Roth, W.-M. (2005b). *Doing qualitative research: praxis of methods*. Rotterdam: Sense Publishers.

Roth, W.-M. (2007). Emotion at work: a contribution to third-generation cultural historical activity theory. *Mind, Culture and Activity*, *14*, 40–63.

Roth, W.-M. (2010). An anthropology of reading science texts in online media. *Semiotica*, *182*, 409–442.

Roth, W.-M. (2012). *First-person method: for a rigorous approach to the study of living/lived experience*. Rotterdam: Sense Publishers.

Varela, F. J. (2001). Intimate distances: fragments for a phenomenology of organ transplantation. *Journal of Consciousness Studies*, *8*, 259–271.

Vidal-Madjar, A. (1978). Teaching mathematics and statistics to adults who are keen on psychology. *Educational Studies in Mathematics*, *8*, 381–390.

Vygotsky, L. S. (1927/1997). The historical meaning of the crisis in psychology: a methodological investigation. In W. R. Rieber & J. Wollock (Eds.), *The collected work of L. S. Vygotsky* (Vol. 6, pp. 233–343). New York: Kluwer Academic.

Wittgenstein, L. (1976). *Wittgenstein's lectures on the foundations of mathematics, Cambridge, 1939*. Hassocks: The Harvester Press.

Wittgenstein, L. (1997). *Philosophische Untersuchungen [Philosophical investigations]* (2nd ed.). Oxford: Blackwell Sci. (First published in 1953).

Wagner, D. (2011). Opening mathematics texts: resisting the seduction. *Educational Studies in Mathematics*. doi:10.1007/s10649-011-9372-8.

Wagner, D., & Davis, B. (2010). Felling number: grounding number sense in a sense of quantity. *Educational Studies in Mathematics*, *74*, 39–51.

Commentary on the Chapters on Probability from a Stochastic Perspective

J. Michael Shaughnessy

1 Introduction

The predominant probability perspective of authors in this section of the book arises from the oft ill-defined constructs "stochastic" and "random." The four chapters authored by Batanero, Arteaga, Serrano, and Ruiz, Saldanha and Liu, Jolfaee, Zaskis, and Sinclair, and Prodromou all rely on the meanings of these two terms from various points of view and for various purposes. Discussions of 'stochastic' and 'random' are presented in these four chapters with reference to historical, developmental, and modeling perspectives, as these authors present research results on both teachers' and students' conceptions and interpretations of randomness. The chapters by English and Roth have a different focus than these four. English argues that young children not only can but should have experiences with constructing their own structures and representations of data. Roth's thesis is that a Bayesian approach to probability should be included in introductory statistics courses. While English's work with young children provides a basis from which to build toward the development of future connections between probability and statistics, the Roth chapter has little if any connection to a stochastic perspective of probability. Prior to further reflection on the individual chapters in this section, I want to attempt some synthesis of such terms as stochastic, random, and probability from this collection of authors to provide a lens for reflecting on their contributions.

2 Stochastic

The word 'stochastic' is regarded as an adjective in dictionaries. Among its meanings are: *random; involving probability; involving guesswork*. In mathematics and

J.M. Shaughnessy (✉)
2820 SW Huber St, Portland, OR 97219, USA
e-mail: mikesh@pdx.edu

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_25,

statistics, one refers to 'stochastic processes', which involve repeatable trials and possibly branching systems where each branch is assigned a probability for a path leading to a subsequent trial. It is interesting that one of the meanings of 'stochastic' is 'random', for these terms can become circular in their meaning or in our everyday language. In the teaching and learning of probability and statistics, the word 'stochastics' has come into popular use, especially in Europe. The word has now become a noun used by both psychologists and educators in statistics education research and in curriculum development. Generally, the word 'stochastics' is now often used as shorthand for 'probability and statistics.' However, with such a generalization of the word we can lose its original meaning.

Saldanha points to the distinction made by Liu and Thompson (2007) between a stochastic and a non-stochastic conception of events. Stochastic events involve an underlying repeatable process, as opposed to a one time, non-repeatable event. Pulling samples of people from a population (say, a city) to measure some trait (say, foot length) is a repeatable process. Tossing a thumbtack is a repeatable process. Having pancakes for breakfast on May 15, 2013 is a one-time event; it is not a repeatable process, not for that one day. A stochastic perspective on probability involves repeatable processes, such as repeated Bernoulli trials of a probability experiment, or pulling repeated samples to generate a sampling distribution of some measure. Such a perspective leads rather naturally to thinking about probability in terms of long-range relative frequency of the outcomes of a repeatable process, and therefore also to forming very close connections between probability and statistics, for one can generate distributions of outcomes from repeatable processes, empirical sampling distributions, for example. Distributions in turn contain lead to some fundamental concepts in statistics such as measures of shape, center, and variation of distributions. Therefore, while the term 'stochastic' does indeed encompass both probability and statistics, there are also limitations to what it encompasses in the realm of probability, for there are assumptions included in what we mean by repeatable. Repeatable processes normally are assumed to be independent from one trial to the next. In the realm of mathematical probability, the word 'independent' means that the probability of an outcome does not change from one trial to the next. Furthermore, we ordinarily expect some variability in the outcomes of the repeatable process; else it would not be interesting to study the accumulation of outcomes. After all, variation is the cornerstone of statistics, the very reason that the discipline of statistics came into existence.

Thus far our discussion of 'stochastic' has led us to the following:

A stochastic perspective on probability involves:

1. A repeatable process
2. Repetitions that are statistically independent
3. Variability in the outcomes of the repetitions

If trials in some repeated process were dependent, if the probabilities might change after each trial, we would need to update the system based on new information to compute new probabilities after each trial. In this context, probability might better be modeled from a Bayesian perspective, as in the chapter by Roth.

I agree with Roth that a Bayesian approach to probability is important for students to learn about in an introductory statistics class, however, there is nothing particularly stochastic about the Bayesian approach. It is entirely based on revising proportions based on additional information, either prior or posterior information. There is nothing easily repeatable in the Bayesian model of probability.

Modeling is yet another concept that can help us to further pin down the meaning of stochastic. Many of the chapters in this section point to the various attempts to model randomness in the historical emergence and development of probability. These include Laplace's classical equiprobable set of outcomes in a finite sample space, Von Mises (1952) long range relative frequency, Kolmogorov's (1950) measure theory approach to providing some axiomatics for probability, the Bayesian approach to probability, and de Finetti's (1970) subjective probability—these can all be thought of as attempts to *model* chance. A stochastic perspective on probability involves both empirical and theoretical probability distributions and modeling the interaction between them, as emphasized in the chapters by Saldanha and Liu, and Prodromou.

We might assume a theoretical probability distribution, or even use technological tools such as *Tinkerplots* (Tinkerplots 2012) to create a hypothetical working distribution, and then generate repeated empirical outcomes based on that assumed distribution. Or we could do the reverse—build up from an empirical distribution to a proposed theoretical distribution by simulating repeated trials of an experiment. If we know parameters of a distribution (e.g., population proportion, mean, variation) we can generate empirical result that we expect to mirror the known population. Conversely (and more often the goal of statistics) when we do not know parameters of a parent population, information from repeated samples can provide us with tools to create an empirical distribution to estimate unknown aspects of the theoretical population distribution. This back and forth process between empirical and theoretical distributions involves modeling probability from a stochastic perspective. It involves model building, model eliciting, and model evaluating activity from students who are engaged in curriculum activities designed from a stochastic perspective on probability, a long range relative frequency approach to probability.

In summary, a stochastic perspective to probability involves:

1. Repeatable processes
2. Repetitions that are statistically independent
3. Variability in the outcomes of the repetitions, and
4. Modeling probability in terms of long range relative frequency

Let us be mindful that this stochastic perspective is only one model among several approaches to probability. As Saldanha and Liu remind us in their chapter:

> … the meaning of probability one employs in a particular situation should be determined by the tasks and the types of problems that are involved. From this perspective, designing a probability curriculum and instruction is no longer a task of deciding which side to take, classical probability or modern axiomatic probability, subjective or objective, and so on. Rather, instructional design becomes an enterprise in which one considers the strengths and weaknesses of each theory, and elicits useful constructs and ways of thinking in designing the instructional objectives.

3 Random

Meanings assigned to the word ‘random’ and interpretations of that term are even more diverse than the word stochastic. The chapters by Batanero et al. and Jolfaee et al. are particularly concerned with teachers and students (respectively) conceptions of randomness; and most of the other chapters introduce the word random at some point. A dictionary can provide a lengthy list of meanings for the word random, including: *chance, accidental, haphazard, arbitrary, causal, unsystematic, indiscriminant, unplanned*, and *unintentional*. Most of these meanings refer to everyday colloquial use of random, only *chance* and *causal* suggest possible reference to a more statistical or probabilistic perspective on the word random. According to several of the authors in this section, a statistical perspective on randomness must include notions of a distribution of possible outcomes; else we would be confined to thinking of random only as unpredictable, in the colloquial sense.

The chapters by Batanero et al., Jolfaee et al., and Saldanha and Liu each begin with a re-iteration of the various historical attempts to model randomness—Classical, Frequentist, Bayesian, Subjectivist—followed by a tour of past research on peoples’ conceptions of randomness, particularly conceptions about random sequences. In this regard, it is unfortunate that the chapter authors didn’t have an opportunity to work together ahead of time in providing the research and theoretical background for their chapters, as they all repeat nearly the exact same historical overview.

Research on peoples’ conceptions of randomness has been primarily based on their ideas and beliefs about randomly generated sequences. In this regard, the word ‘random’ is really only defined in retrospect, looking back at some outcomes and asking people to decide which are random, and which are not, and why. Attempts to define random can involve probability, and easily become circular, because random involves probability, yet probability involves random. It is tricky terrain. Piaget and Inhelder (1951/1975), Green (1989), Fischbein (1975), and Batanero and Serrano (1999) among others, have contributed to the development of the literature on students’ conceptions of randomness. Research on random sequences has produced robust results that have repeatedly confirmed that subjects are:

1. Not very adept at generating random sequences themselves
2. Not very capable of distinguishing randomly generated sequences from non-randomly generated ones

For example, people tend to expect shorter runs in random binomial sequences than actually occur in practice, and to overestimate the number of switches in a sequence. Recent attempts to define a random sequence have been based on statistical tests or on complexity theory. Bennett (1998) proposed to define random in terms of the number of statistical tests that it satisfies. Li and Vitany (2008) refer to work concerning the length of the algorithm necessary to describe a sequence, a complexity theory approach. The complexity theory approach is an updated version of Von Mises (1952) intuitive description of a random sequence as impossible to forecast.

Thinking of probability in terms of random sequences seems quite limiting, and perhaps does not provide us with the best lens to investigate student thinking about probability, statistics, or things stochastic. The statistical concepts of independence, variability, and distribution are more easily defined than a term like 'random', and perhaps provide more traction for us to investigate students' thinking and learning about probability. Although we can have interesting philosophical and epistemological discussions about 'random' or 'randomness', I do not believe it is in either our own or our students' best interest to spend more time researching random sequences. It's pretty much a done deal. Humans are clearly not good at either generating or identifying random sequences. Providing them with some information as to why they are not good at it, and what random sequences really do look like, is probably a worthwhile endeavor. But overall, random sequences are not and should not be in the main stream of the stochastic education of our students. Better we emphasize variation, independence, and distribution as the major foci of both research and practice. As I understand it, that is more in line with the recommendations in the chapters by Saldanha and Liu, and Prodromou.

4 Further Reflections on Individual Chapters

Given this overview of some common elements and themes about the stochastic perspective that pulse throughout this collection of chapters, I'm in a position to comment in a bit more detail on the chapters.

English's thesis that statistical literacy must begin in the earliest years of schooling may be a radical concept in some countries, but it is a position that is both responsible and defensible. The research conducted by English is defensible because it adds to our previous knowledge of young children's ability to begin to understand both probability concepts and data representations (e.g., Fischbein 1975; Jones et al. 1999, 2000). English presents evidence for the creativity and flexibility that young children demonstrate when they are set loose to create their own structures and representations of data. Her research flies in the face of current beliefs of curriculum developers with regard to what is developmentally appropriate in the stochastic realm for young children. English's position is also responsible because she encourages us to plant the seeds for the progression of statistical literacy in schools early on. Statistics education needs to ramp up throughout the early school years so that by the time students are in those middle and secondary years they have some robust conceptions to build upon. Perhaps students who had experienced an early start to in developing their statistical literacy would respond differently than the students whose thinking is described in this section in the studies of Jolfaee et al., Saldanha and Liu, and Prodromou. The neglect of an adequate elementary school statistics progression is one of the most glaring weaknesses in the first version of the newly adopted Common Core State Standards for Mathematics in the United States (2010). The Common Core Standards claim to be research based, but evidently not all content areas in the CCSSM were equally grounded in research on student learning.

Batanero et al. confirm that pre-service elementary teachers' perceptions of randomness, as manifested in characteristics of random sequences, are quite similar to those found by previous researchers among students (Green 1983; Batanero and Serrano 1999). Teachers' decisions on which sequences were random were based on faulty perceptions about the number of switches, the length of runs in the sequence, or combinations of the two. This is not at all surprising because pre-service elementary teachers tend to have no more background and experience with probability and statistics than members of the general population. The implication, of course, is that such teachers will not be in an adequate position to support the development of statistical literacy in their own students without considerable experience and training in probability and statistics. Without sufficient experience of their own teachers will be doomed to perpetuate the cycle of statistical illiteracy that permeates our schools for yet another generation. One wonders how the teachers in the Batanero et al. study could provide the opportunity and support for their own students to investigate the data sets used in the English study. Not very well at all, I would think.

Once again I raise the question as to whether 'random' is the right focus for teaching our students about the stochastic perspective of probability. There's more to randomness than just sequences. For example, could students tell the difference between real and fake distributions of outcomes from a probability experiment (Shaughnessy et al. 2004)? What tools would they use to identify data generated from a probability experiment versus data that was faked? What would be reasonable to expect about variation across sample results in a sampling distribution? The number of runs in random sequences involves variation, and the length of runs involves the concept of independence. Perhaps variation and independence are better starting points for teachers and students to develop and investigate statistical literacy than random sequences. As we have seen in our earlier discussion, randomness can easily become an epistemological conundrum.

Jolfaee et al. investigate middle school students' conceptions of randomness, along with the gestures the students use when engaged in discourse about random events. The students were asked to provide their own examples of random, and as one might expect most of the students produced examples that involved unpredictability or unexpected results. For some students, random meant that there was variability, that things won't be the same every time. Of note in the Jolfaee et al. work is that none of the student generated examples of randomness involved producing sequences of random outcomes. When left to construct their own examples, rather than evaluate the randomness of researcher generated examples of random sequences, students provided examples from everyday life, colloquial, non-statistical uses of the term 'random.' One question I have for Jolfaee et al.: What did the accompanying work on analyzing student gestures add to your analysis of students' thinking about randomness? Did you obtain significantly different information about student thinking by including the gesture piece than you would have without it? What, if anything, did the work on gesture add to your study? It is not clear in this chapter whether there was any value-added from the gesture component of the study.

As mentioned previously, the chapter by Roth does not fit within the stochastic perspective because it does not involve thinking about probability as arising from a repeatable process. Conditional probability perspectives take the point of view that we should update probability estimates when we obtain new information. The problems that are attacked from a Bayesian perspective are much concerned with events that are dependent, as opposed to the repeatable, independent events in the stochastic perspective to probability. That said, Roth has set forth a very important thesis that students in introductory statistics courses should experience a Bayesian approach to calculating probabilities. Given the research by psychologists on heuristics that people appeal to when estimating the likelihood of events (e.g., Kahneman and Tversky 1972; Gigerenzer 1994), and the difficulty that students have with understanding conditional probability and the use of contingency tables in decision-making Batanero et al. (1996), I most certainly agree that Roth has a point.

However, in my view Roth does not provide a strong case for teaching the Bayesian approach in this chapter. If anything, this microanalysis of his own personal discovery of the Bayesian approach clouds the real issue of why a Bayesian approach is important to include in introductory statistics. There are problems that a Bayesian approach can best model, and Roth provides little if any support for solving Bayesian problems in this introspective case study report. The chapter rambles at times more like a personal blog than a research paper. Roth includes a diatribe in defense of case based research methodology, along with some totally unsupported claims that the constructivist approach is fallacious in its thinking. Like any methodology, case-based research has its assumptions, strengths, and limitations, and is certainly a viable methodology to investigate certain questions. It may even have been an appropriate methodology to investigate the viability of including Bayesian approaches in introductory statistics, but not as presented in this chapter. One of the biggest problems with this personal case is that the reader has no context for the naked proportions that suddenly appear in the author's retrospective reconstruction of how he began to appreciate Bayesian probability. Where did 17/20 and 6/20 come from? What was the original problem that was being represented via conditional probabilities? If one is making a case for the importance of teaching a topic, the relevance of that topic, the context and possible applications and types of problems that Bayesian calculations can address need to be part of the conceptual analysis that supports the importance of teaching that topic. It is not clear that Roth makes a case in this chapter for the efficacy of doing Bayesian statistics with adult learners. Quite the contrary, reading this chapter may have the opposite effect altogether, and drive potential instructors of a Bayesian approach away altogether.

The two chapters, by Saldanha and Liu, and by Prodromou, have a common theme that the teaching and learning of probability involves modeling chance. Probability as long range relative frequency can be modeled by using simulations of repeated experiments—stochastic processes—that for Saldanha and Liu distinguish between a distributional mode and a single event mode in thinking about probability. Both these chapters argue that the conceptual interaction between theoretical probability distributions and empirically generated sampling distributions (often created via technology) should be the lynchpin for the teaching and learning of probability and statistics.

For Prodromou, this interaction can provide an avenue to connect back to the classical conception of probability, a link between distributions of data and theoretical probability distributions. She begins her chapter with a lament about the lack of instruction in probability in schools today, and how statistics is taking over the school curriculum due to the easy access to technology. For Prodromou, a modeling approach can help to synthesize the two (sometimes) conflicting approaches to probability, the de Moivre–Laplace classical sample space model of probability and the statisticians' view of probability as long term relative frequency. Otherwise, Prodromou would agree with Borovcnick (2011) that the probability curriculum will remain isolated in a world of dice, spinners, and coins, and gradually vanish altogether due to its irrelevance to the real world. Students in Prodromou's research create models of theoretical probability distributions using the *Tinkerplots* software (Tinkerplots 2012), and then generate sampling distributions of results using the Sampler tool in *Tinkerplots*. They investigate the interaction between the results shown in their data and an initial starting distribution, looking for expected similarities (shape, center, spread, etc.) along the way. *Tinkerplots* allows students to set up populations with multiple traits and investigate complex multivariate worlds. Multivariate data sets provide a rich context for students to explore proportions of outcomes with various characteristics, and to compare the data to the original model that was created in the *Tinkerplots* mixer.

Likewise, Saldanha and Liu build a case for a stochastic approach to probability that involves thinking of events as the result of some repeatable process that produces collections of events that subsequently lead to modeling probability as long term relative frequency. In this regard they state, "We concurred with a number of researchers' thesis (Konold 1989; Shaughnessy 1992) that when designing instruction probabilistic reasoning is better thought of as creating models of phenomena." Thinking about probability from a modeling perspective can provide considerable flexibility in our teaching and instructional materials. One could model probability based on the types of questions and problems that we address, rather than in a one size fits all approach to probability. A modeling context for probability encourages the inclusion of Bayesian, classical, and frequentist approaches to probability because each of these perspectives is helpful in re-solving different types of probability problems.

In summary, I had two principal takeaways from this collection of chapters. One conclusion from reading the work of these authors is that while the stochastic perspective is widely applicable, the world of probability must include more than just the stochastic perspective for our students. Another conclusion is that conceptualizing probability in terms of models and modeling is a promising perspective on probability for both future researchers and future teachers of probability. A modeling perspective leads to thinking of probability in several ways, depending on the problem being investigated, and it provides a supportive environment for the interaction between empirical probability distributions and theoretical probability models. Flexibility should be prominent in our thinking about the meaning of probability, both in our research stance, and for teaching our students at all levels.

References

Batanero, C., & Serrano, L. (1999). The meaning of randomness for secondary school students. *Journal for Research in Mathematics Education*, *30*(5), 558–567.

Batanero, C., Estepa, A., Godino, J. D., & Green, D. R. (1996). Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education*, *27*, 151–169.

Bennett, D. J. (1998). *Randomness*. Boston: Harvard University Press.

Borovcnick, M. (2011). Strengthening the role of probability within statistics curricula. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics—challenges for teaching and teacher education. A joint ICMI/IASE study* (pp. 71–83). New York: Springer.

Common Core State Standards for Mathematics (2010). http://www.corestandards.org/the-standards/mathematics.

de Finetti, B. (1970). *Theory of probability: a critical introductory treatment*. London: Wiley.

Fischbein, E. (1975). *The intuitive sources of probabilistic thinking in children*. Dordrecht: Reidel.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is relevant for psychology (and vice versa). In G. Wright & P. Ayton (Eds.), *Subjective probability*. New York: Wiley.

Green, D. R. (1983). A survey of probabilistic concepts in 3000 students aged 11–16 years. In D. R. Grey et al. (Eds.), *Proceedings of the first international conference on teaching statistics* (Vol. 2, pp. 766–783). Sheffield: University of Sheffield.

Green, D. R. (1989). Schools students' understanding of randomness. In R. Morris (Ed.), *Studies in mathematics education: Vol. 7. The teaching of statistics* (pp. 27–39). Paris: Unesco.

Jones, G. A., Langrall, C. W., Thornton, C. A., & Mogill, A. T. (1999). Students' probabilistic thinking in instruction. *Journal for Research in Mathematics Education*, *30*, 487–519.

Jones, G. A., Thornton, C. A., Langrall, C. W., Mooney, E. S., Perry, B., & Putt, I. J. (2000). A framework for characterizing students' statistical thinking. *Mathematical Thinking and Learning*, *2*, 269–307.

Kahneman, D., & Tversky, A. (1972). Subjective probability: a judgment of representativeness. *Cognitive Psychology*, *3*, 430–454.

Kolmogorov, A. N. (1950). *Foundations of the theory of probability*. New York: Chelsea.

Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction*, *6*(1), 59–98.

Li, M., & Vitany, P. (2008). *An introduction to Kolmogorov complexity and it applications*. New York: Springer.

Liu, Y., & Thompson, P. W. (2007). Teachers' understandings of probability. *Cognition and Instruction*, *25*(2), 113–160.

Piaget, J., & Inhelder, B. (1951/1975). *The origin of the idea of chance in children*. London: Routledge & Kagan Paul.

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.

Shaughnessy, J. M., Ciancetta, M., Best, K., & Canada, D. (2004). *Students' attention to variability when comparing distributions*. Paper presented at the research presession of the 82nd annual meeting of the National Council of Teachers of Mathematics, Philadelphia, PA.

TinkerPlots (2012). *Dynamic data exploration (Version 2.0) [Computer software]*. Emeryville: Key Curriculum Press.

Von Mises, R. (1952). *Probability, statistics and truth*. London: William Hodge and Company. [Original work published in 1928].

Perspective IV: Mathematics Education

Preface to Perspective IV: Mathematics Education

Egan J. Chernoff and Gale L. Russell

Two main threads are beginning to emerge within research investigating probabilistic thinking in the field of mathematics education. Although it is premature to declare, with certainty, the genesis of these threads, particular possible factors (to name a few) include: probability's recent entry into curricula worldwide, the age of the field, the volume of research being produced. Regardless of the origins, what is clear is that the two threads in the research investigating probabilistic thinking are: (content) topics in probability (e.g., sample space) and areas of research in mathematics education (e.g., affective domain). (Worthy of note, the two threads are not mutually exclusive.)

Research investigating (content) topics in probability has utilized and resulted in a variety of theories, frameworks, and models, which are used to account for the development of students' (and teachers') probabilistic understanding, reasoning and, in more general terms, thinking (see, for example, Mooney, Langrall and Hertel, this volume). Specific topics in probability examined in this volume include: binomial probability (Sanchez and Landin), conditional probability (Huerta), counting methods (Maher and Ahluwalia), sample space (Mamolo and Zazkis; Paparistodemou) and others. In the investigation of these topics, specific frameworks are utilized and developed in this volume (Mooney, Langrall and Hertel; Prediger and Schnell; Sanchez and Landin). Thus, the research housed in this volume not only contributes to research on particular topics in probability, but, further, to the theories, models, and frameworks utilized and developed in the research literature. As a result, this thread contributes, through particular topics in probability, to research investigating probabilistic thinking in mathematics education.

In addition to the thread of research investigating particular topics in probability, probabilistic investigations, linked to specific areas of research in mathematics education, are also housed in this volume. Specific examples found in this perspective include: the affective domain (Nisbet and Williams), culture (Sharma), teaching

E.J. Chernoff (✉) · G.L. Russell
University of Saskatchewan, Saskatoon, Saskatchewan, Canada
e-mail: egan.chernoff@usask.ca

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_26,

experiments (Nilsson; Prediger and Schnell), and teachers' probabilistic knowledge and context (Mamolo and Zazkis). Thus, this thread of research into probabilistic thinking may disseminate to and inform different areas of research in mathematics education and, eventually, the larger domain.

The research found in this, the mathematics education perspective, as detailed above, demonstrates a cross-fertilization between research in probabilistic thinking and research in mathematics education. More specifically, in one thread, research of particular topics in probability, which utilizes theories, frameworks, and models from mathematics education, results in both the creation of new and refinement of existing theories, frameworks, and models. For the other thread, research investigating probabilistic thinking contributes to other areas of research in mathematics education. Perhaps the future existence of an established cross-fertilization *within* the field of mathematics education research will result in cross-fertilization *beyond* mathematics education.

A Practitional Perspective on Probabilistic Thinking Models and Frameworks

Edward S. Mooney, Cynthia W. Langrall, and Joshua T. Hertel

Abstract Our goal in this chapter is to showcase the various theories and cognitive models that have been developed by mathematics and statistics educators regarding the development of probabilistic thinking from a "practitional" perspective. This perspective juxtaposes the work of mathematics and statistics educators with research on probability from the field of psychology. Next, we synthesize theories and models that have been developed for specific probability concepts and processes. These models suggest different levels or patterns of growth in probabilistic reasoning. Finally, we examine the role these models have had on instruction and curriculum recommendations.

> *Mathematics education should identify its specific problems and should try to build its own concepts and its own research strategies. Psychology certainly may be very helpful but only as a general theoretical framework.*
>
> Fischbein (1999, p. 53)

More than any other topic in mathematics education, the study of probability has been rooted in the field of psychology. The previous chapter showcased some of the prominent psychological contributions to the research literature and presented recent developments in psychology that appear promising to future probability research. In this chapter, we shift the focus to the theories and cognitive frameworks characterizing the development of probabilistic thinking from the perspective of mathematics education. First, we describe what we have coined the *practitional* perspective, as a means to juxtapose the work of mathematics and statistics educators studying probability with that of psychologists. Next we summarize the research literature on probabilistic thinking frameworks, which suggest different levels or patterns of growth in probabilistic reasoning that result from maturational or interactionist effects in both structured and unstructured learning environments. Finally, we discuss the potential and influence of these frameworks from a practitional perspective.

E.S. Mooney (✉) · C.W. Langrall · J.T. Hertel
Normal, IL, USA
e-mail: mooney@ilstu.edu

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_27,

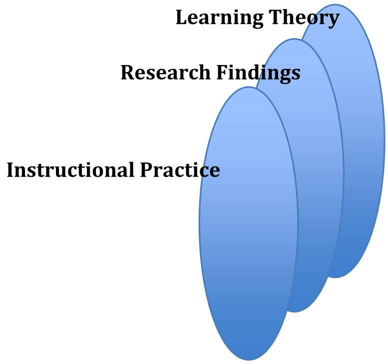


Fig. 1 Facets of a practitional perspective

1 What Is a "Practitional" Perspective?

We use the word practitional to distinguish the general interests and concerns of mathematics educators from those of psychologists. Following Shaughnessy (1992), we consider the psychological perspective as one of *observing* and *describing*, whereas the practitional perspective is aimed at *intervening*. The lens of a psychologist is broadly directed at human learning and thought; the practitional lens of a mathematics educator takes a sharper focus on improving teaching practice and increasing student learning. Moreover, the practitional lens is multifaceted; it is centered on instructional practice and informed by findings from mathematics education research and theoretical positions on learning mathematics (see Fig. 1).

Much of the work of mathematics educators is grounded in practice. Thus, in taking a practitional perspective, we shift focus from the psychological foundations that undergird research on probability to a more educationally-oriented perspective that recognizes mathematics education as drawing from multiple sources of knowledge including classroom practice and learning theory. This affords us a new lens through which to present and interpret probabilistic thinking models and frameworks in terms of what they mean for the teaching and learning of probability. Moreover, we believe that the practitional perspective allows us to address issues and ideas that go beyond those typically considered by psychologists whose research is conducted within the domain of mathematics, but not necessarily within the culture of mathematics education. This stance reflects our response to the quote by Fischbein (1999), with which we opened the chapter.

2 Why Take a Practitional Perspective?

We draw our inspiration for the practitional perspective largely from the work of Efraim Fischbein, who saw the need for mathematics education researchers to do more than simply use the theories of cognitive psychology. In the article, *Psychology and Mathematics Education*, Fischbein (1999) discussed the rift between psychology and mathematics education. He noted, "despite researchers' attempts to apply

Piaget's ideas to mathematics education, his [Piaget's] work did not constitute a bridge between psychology and mathematics education" (p. 48). Fischbein argued that Piaget was not concerned with the influence of instruction on the development of reasoning (mathematical or otherwise). Consequently, questions concerning the effects of instruction or school experience were not of interest to Piaget; however, these issues are of interest to education researchers. In particular, knowing how instruction might influence student understanding is a centerpiece of much educational research. From our practitional perspective, Fischbein's argument highlights the need for considerations of teaching and instructional practice. As teachers, we are always faced with the question of student prior knowledge and this greatly influences our decisions in the classroom. As mathematics education researchers, we are faced with the challenge of improving the teaching and learning of probability within a classroom context, not an isolated, research setting. This suggests that our approach should be one that is mindful of lessons from practice.

A second argument put forward by Fischbein was that Piaget's belief in the logical nature of intellectual development gave little credence to unconscious and conscious structures. In other words, Fischbein questioned Piaget's description of the developmental progression. A substantial part of Fischbein's research agenda was in response to what he perceived as a "basic deficiency in the Piagetian doctrine" (Fischbein 1999, p. 49). Specifically, Fischbein took issue with the way that Inhelder and Piaget (1958) explained children's intuitive thinking using logical structures. He noted,

> If one investigates the student's difficulties and misconceptions, one does not identify only logical deficiencies. One identifies, very often, intuitive tendencies, intuitive interpretations, and models—tacit or conscious—that contradict the formal knowledge with which school tries to endow the student. Without this picture, one knows only a part of the truth with regard to the mechanisms of the student's difficulties and successes. (p. 49)

Helping students overcome "intuitive tendencies, intuitive interpretations, and models" is clearly a goal of mathematics educators. As *interveners* rather than *observers* we seek to understand what a student knows about a topic before instruction and recognize that this prior knowledge, intuitive or otherwise, plays a critical role in student understanding.

Fischbein's objection led him to investigate intuitive thinking in mathematics and science. His work, which spanned various topics in mathematics, led to two classification schemes for intuition. The first scheme categorizes intuitions by considering the relationship between an intuition and a solution to a problem. Using this scheme, an intuition falls into one of four categories: affirmatory, conceptual, anticipatory, or conclusive (for a more detailed description of the first categorization, see Fischbein 1987). The second classification system considers whether or not a particular intuition developed within the context of instruction. Fischbein classified intuitions that developed independently of any systematic instruction as *primary intuitions*. These intuitions, he explained, were a result of one's personal experience. Intuitions resulting from instruction were classified as *secondary intuitions*. Fischbein (1987) noted that development of secondary intuitions relied on participation by the learner in structured activity,

> Such a process implies, in our view, the personal involvement of the learner in an activity in which the respective cognitions play the role of necessary, anticipatory and, afterwards, confirmed representations. One may learn about irrational numbers without getting a deep intuitive insight of what the concept of irrational number represents. Only through a practical activity of measuring may one discover the meaning of incommensurability and the role and meaning of irrational numbers. (p. 202)

Fischbein stressed that secondary intuitions are formed when an individual is involved in an activity that requires serious consideration of a particular idea. Thus, his second classification system, which was built upon psychological research, learning theory, and knowledge of practice, represents the balance of viewpoints embodied by the practitional perspective.

It is also significant to our thesis that Fischbein's early research was conducted within the domain of probability. He examined students' intuitions about probability in instructional contexts (e.g., Fischbein and Gazit 1984; Fischbein et al. 1970) and characterized their understanding according to developmental stages that reflected the influence of instruction (Fischbein 1975). As Greer (2001) noted, "the interplay between intuitions, logical thinking, and instruction is central to Fischbein's theory of the development of probabilistic thinking and makes it simultaneously a theory of instruction" (p. 19).

3 Frameworks Characterizing the Development of Probabilistic Thinking

Much has been written about the development of probabilistic thinking, and we direct the reader to the following comprehensive reviews: Hawkins and Kapadia (1984); Shaughnessy (1992); Borovcnik and Peard (1996); Jones (2005); and Jones et al. (2007). Collectively, these works trace the progress of research from Piaget's seminal work on children's development (Piaget and Inhelder 1975) to Tversky and Kahneman's (1974; Kahneman and Tversky 1972, 1973) groundbreaking studies on judgment heuristics and biases, to more recent studies about probabilistic thinking on a variety of related constructs. Because an exhaustive summary of the probability models and frameworks that have been reported in the literature is beyond the scope of this chapter, we have restricted our focus to a select sample of frameworks that illustrate the multifaceted nature of probabilistic thinking. We summarize the key components of these frameworks, synthesize characterizations of students' probabilistic thinking across the frameworks, and consider their contribution to the field from a practitional perspective.

3.1 SOLO Taxonomy

Many prominent frameworks characterizing the development of probabilistic thinking have been based on a neo-Piagetian model of cognitive development, the Structure of the Observed Learning Outcomes (SOLO) taxonomy (Biggs and Collis 1982, 1991). The SOLO taxonomy incorporates five modes of functioning: sensorimotor (from birth), ikonic (from around 18 months), concrete symbolic (from around 6 years), formal (from around 14 years), and post formal (from around 20 years). The emergence and development of these modes of functioning is such that earlier modes are incorporated into later ones. Furthermore, within each mode, learners cycle through three cognitive levels—unistructural, multistructural, and relational—whereby each level subsumes the preceding one and represents a shift in the complexity of students' thinking. At the unistructural level of thinking, the learner focuses on a single relevant aspect of a task or idea. A learner at the multistructural level recognizes several aspects of a task or idea but is unable to consolidate them. At the relational level, the learner can integrate the aspects of the task or idea in a coherent manner. There are also two levels that characterize thinking that shifts between modes; at the prestructural level a learner functions in ways characteristics of the previous mode and at the extended abstract level a learner functions in ways characteristic of the next mode.

3.2 Prominent Probabilistic Thinking Frameworks

We have identified the most prominent probability frameworks in the research literature, all of which use the SOLO taxonomy as a basis for characterizing the development of students' thinking. Two research groups are responsible for the development of these frameworks; one in the United States led by Graham Jones (Jones et al. 1997; Tarr and Jones 1997) and another in Australia led by Jane Watson (Watson et al. 1997; Watson and Caney 2005; Watson and Moritz 2003). An overview of the key components of these frameworks is presented in Fig. 2.

These frameworks were developed and validated through research conducted with school-age students; they illustrate thinking generally indicative of the concrete-symbolic mode as described in the SOLO taxonomy. The ikonic mode of functioning was exhibited by students in all of the framework development studies and is reflected in the inclusion of the prestructural level (subjective, prestructural/ikonic) in each of the frameworks. It is noteworthy that researchers have not characterized thinking at the extended abstract level, which would bridge into the formal mode of functioning. To our knowledge, the development of this level of sophistication of probabilistic thinking has not yet been characterized in an organized framework, such as the ones presented here.

The frameworks described in Fig. 2 show a strong consistency across a wide range of students and cultures. Collectively, they illustrate the complexity of probabilistic thinking and address fundamental ideas such as fairness, luck, chance, randomness, and likelihood. The frameworks also examine constructs for sample space, probabilities of simple and compound events, probability comparisons, conditional probabilities, and independence. These frameworks organize a wealth of research-based knowledge about student thinking—the nature of its development and the intuitions and misconceptions students bring to instruction.

	Probability concepts	Data sources	Organization of cognitive levels	Additional notes
Jones et al. (1997)	• Sample space • Probability of an event • Probability comparisons • Conditional probability • Independence	Case study with Grade 3 students in U.S. (teaching experiment, individual interviews)	Subjective Transitional Informal quantitative Numerical	Validated with students in Russia (Volkova 2003) Extended and validated with students in Lesotho (Polaki et al. 2000)
Tarr and Jones (1997)	• Conditional probability • Independence	Case study with Grades 4–8 students in U.S. (individual interviews)	Subjective Transitional Informal quantitative Numerical	Extended constructs from Jones et al. framework
Watson et al. (1997)	Chance measurements (simple events, likelihood, comparison of events)	Written survey with Grades 3, 6, 9 students in Australia	Ikonic Unistructural (U) Multistructural (M) Relational (R)	Identified two cycles U_1-M_1-R_1 $\rightarrow$ U_2-M_2-R_2 whereby the 2nd cycle shows consolidation in more complex setting Confirmed with longitudinal data (Watson and Moritz 1998)
Watson and Moritz (2003)	Fairness	Individual interviews with Grades 3, 5, 6, 7, 9 students in Australia	Ikonic Unistructural Multistructural Relational	
Watson and Caney (2005)	Random Processes (meaning of random, chance language, luck, equal-likelihood)	Written survey and interviews with Grades 3, 6, 9 students in Australia	Ikonic Unistructural Multistructural Relational	

Fig. 2 Key components of prominent probability frameworks

We cannot do justice to the richness of these frameworks within this chapter and refer the reader to the publications cited for detailed discussion and information about their development. However, we have attempted to synthesize key characteristics across the frameworks in order to present an overall picture of the development of school students' probabilistic thinking (see Fig. 3).

Our synthesis shows that, in general, students' probabilistic thinking moves from being idiosyncratic to proportional in nature. Prestructural probabilistic thinking can best be described as subjective. Students' responses to probabilistic tasks are often irrelevant, non-mathematical, or personal in nature (Jones et al. 1997; Watson et al. 1997). They consider probability as something that may or may not happen and outcomes of events may be controlled (Tarr and Jones 1997). Thus, at this level of thinking, students have difficulty creating complete sample spaces or distinguishing between fair and unfair probabilistic situations. Reliance on idiosyncratic thinking makes it difficult for students to predict most or least likely events, compare probabilities, determine fairness, address conditional probability or recognize independence (Tarr and Jones 1997; Watson and Moritz 2003). However, despite these obstacles, students at this level of thinking seem to have an intuitive idea about randomness though they are susceptible to the order or balance of outcomes (Watson and Caney 2005).

Within unistructural probabilistic thinking, students begin to transition from subjective thinking to quantitative thinking. However, their quantitative thinking is typically non-proportional in nature and sometimes flawed. Students often revert back to subjective thinking (Tarr and Jones 1997). Students at this level can predict most or least likely events, compare probabilities, and determine conditional probabilities. They can create complete sample spaces for one-stage experiments and some two-stage experiments as well as distinguish between fair and unfair situations (Jones et al. 1997). At this level of thinking, students begin to become aware that some probabilities change in non-replacement situations and whether consecutive events are related or not; but they tend to develop a positive or negative recency orientation (Tarr and Jones 1997). Students have a more qualitative view about randomness—"anything could happen" (Watson and Caney 2005).

A stronger use of quantitative thinking is indicative of multistructural probabilistic thinking. Counts, probabilities, ratios, or odds are used when making judgments about probabilistic situations. Quantities may not always be correct; however, students can use them to make comparisons and determine fairness (Jones et al. 1997; Tarr and Jones 1997). Sample spaces for one- and two-stage experiments are created more systematically (Jones et al. 1997). Students, at this level, can recognize that conditional probabilities change for all events in without-replacement events. They are also aware of independent events in without-replacement events but may revert to using representativeness strategies when dealing with independent trials (Tarr and Jones 1997). Students are able to provide examples of random situations or methods of generating random processes (Watson and Caney 2005).

Relational probabilistic thinking is based on students' ability to reason with probabilities. Students can use or assign numerical probabilities in complex situations—both equally and non-equally likely situations (Jones et al. 1997). Generative strategies are used to determine the sample spaces of two- and three-stage experiments

Prestructural Probabilistic Thinking
Student thinking is irrelevant, non-mathematical, or personalized.

Students exhibiting thinking at this level…

- Have an intuitive understanding of randomness
- Believe outcomes can be controlled or explained
- Believe consecutive events are always related
- Struggle creating complete sample spaces and distinguishing between fair and unfair situations
- Use idiosyncratic thinking when making predictions

Unistructural Probabilistic Thinking
Student thinking is quantitative and non-proportional.

Students exhibiting thinking at this level…

- Have a tendency to revert to subjective probabilistic thinking
- Have a qualitative view of randomness—"anything is possible"
- Make predictions about most or least likely events using quantities
- Create sample spaces for one-stage experiments and some two-stage experiments
- Distinguish between fair and unfair situations
- Compare probabilities
- Determine conditional probabilities
- Become aware that probabilities can change in consecutive events

Multistructural Probabilistic Thinking
Student thinking is quantitative and proportional.

Students exhibiting thinking at this level…

- Make use of ratios, counts, probabilities or odds in judging probabilistic situations
- Create samples spaces for one-stage and two-stage experiments systematically
- Provide examples of random situations or methods of random generation
- Recognize changes in probability and independence in without-replacement events
- Make predictions that sometimes rely on representativeness strategies

Relational Probabilistic Thinking
Student thinking shows an interconnection of probabilistic ideas.

Students exhibiting thinking at this level…

- Reason with probabilities
- Determine probabilities for complex situations including non-equally likely situations
- Generate sample spaces for two-stage and three-stage experiments systematically
- Connect sample spaces and probabilities
- Examine distribution of outcomes without reverting to representativeness strategies
- Show an appreciation of randomness by not being influenced by order or balance of outcomes

Fig. 3 Synthesis of characterizations across frameworks and explicitly connect sample spaces to probabilities (Tarr and Jones 1997). Students can focus on the distribution of outcomes across trials without reverting to representativeness strategies and are less likely to make predictions when outcomes are equally likely (Tarr and Jones 1997). Students at this level have a greater appreciation of randomness and are not influenced by the order or balance of outcomes (Watson and Caney 2005).

Our summary of probability frameworks shows the strength and consistency these frameworks provide in describing the growth of students' probabilistic thinking. By developing the frameworks through work in classrooms and with students, researchers have considered the instructional implications of their work. Consequently, these frameworks provide a bridge between learning theory and instructional practices.

3.3 Implications for Instructional Practice

Although the previously discussed frameworks present cognitive models that characterize the development of students' probabilistic thinking, they were constructed to serve as tools for teaching and learning. Thus, these frameworks have much to contribute to the scope and sequence of probability curriculum and instruction. All of the researchers have discussed implications for instruction related to their respective frameworks. Jones et al. (1999) found that novice teachers could use a probability framework to develop instruction to improve students' probabilistic thinking. However, they noted that the small-group instructional setting was not typical and further research was needed to determine the feasibility of teachers using the framework in regular classroom settings (Jones et al. 1999).

Watson and Moritz (1998) indicated that the SOLO model of probabilistic thinking could be used to assess student thinking and show changes in thinking over time. They noted that teachers need to be able to make students' probabilistic conceptions explicit and learn to recognize different levels of responses (Watson et al. 1997). Teachers need to recognize the limitations in the forms of assessment used and that even tasks that require written justifications can "lead to some responses which are at lower levels than the students are capable of achieving" (Watson et al. 1997, p. 79).

The researchers on probability frameworks even have implications for specific probabilistic ideas. When dealing with randomness, Watson and Caney (2005) recommended that teachers discuss with students the difference between the colloquial and statistical use of the word "random." Watson and Moritz (2003) recommended that teachers provide students with ample opportunities to work with dice in simple, concrete situations prior to using computer-simulated models of dice. These frameworks can make teachers aware of the complexities of students' reasoning about probability and assist them in planning, implementing, and evaluating instruction to foster the development of students' probabilistic reasoning (Jones et al. 1999).

As stated earlier, the developers of these probabilistic frameworks intended for them to inform the way probability is taught in schools. The main question that arises is: What role have these frameworks had in shaping important curriculum recommendations? Can we see the impact in curricula or recommendation documents? Or, is there a striking omission of this systematic body of research-based evidence about how students' learn?

4 The Influence of Probability Frameworks on Curriculum Standards

In response to the questions posed above, we turn our attention to the recently published Common Core State Standards for Mathematics (National Governors Association for Best Practices and Council of Chief State School Officers 2010). The Common Core State Standards for Mathematics (CCSSM) outline learning expectations for K-12 mathematics in the United States. Currently, 90 % of the States have agreed to formally adopt the CCSSM. This has essentially made CCSSM the national standards for mathematics instruction in the United States. In seeking to answer the question "What role have these frameworks had in shaping important curriculum recommendations?", we analyzed the references provided with CCSSM, as well as the standards themselves, to identify evidence of the previously mentioned probabilistic frameworks.

Based upon the sample references provided, it does not appear that CCSSM is rooted in any meaningful way within the probability research tradition that we have outlined. The CCSSM document does not include a full reference list, but rather lists a sample of the works consulted. None of the works in this list is specific to probability although some do feature discussions of probability (e.g., *College Board Course Descriptions*, *Guidelines for Assessment and Instruction in Statistics Education* (*GAISE*) *Report*, *Research Companion to Principles and Standards*). Two references in the list mention statistics. The first is a commentary by Cobb and Moore (1997) on the differences between statistical and mathematical thinking. The second article, Kader (1999), describes a classroom activity aimed at assisting students in understanding variation from the mean. Kader noted that the article "builds a framework for developing students' understanding of the notion of variation from the mean" (p. 398); however, the research basis of this framework is unclear. Thus, the sample reference list included in the CCSSM does not appear to contain any citations to work by mathematics education researchers investigating the teaching and learning of probabilistic concepts (e.g., Watson, Jones). Likewise, none of the probabilistic frameworks mentioned in our review are present within the list. Although these observations may appear to be somewhat arbitrary it is important to note that, in contrast to the absence of probabilistic research, the sample reference list does cite the model of Cognitively Guided Instruction. Additionally, work by researchers in other areas of mathematics education research (e.g., measurement, number) is featured. Our review of the references appears to indicate that existing probability frameworks did not play a prominent role in the writing of the CCSSM. Moreover, as we will argue next, the focus on probability as a topic of study is minimized within the standards.

Reviewing the CCSSM standards themselves, probability is absent from over half of the grade levels. Specifically, standards for Kindergarten, Grade 1, Grade 2, Grade 3, Grade 4, and Grade 5 lack any discussion of probability concepts. In contrast, standards for all of these grade levels feature data analysis. The Grade 6 standard includes probability as a heading, but no learning goals related to probability are listed. In fact, the first instance of probability topics within the CCSSM is in the Grade 7 standard. Although specific probability concepts are listed in the Grade 7 standard, they are again excluded from the Grade 8 Standard. Instead, the Grade 8 standard focuses on "investigating patterns of association in bivariate data" (p. 56). At the high school level, CCSSM includes a Statistics and Probability standard that is divided into four components: (a) interpreting categorical and quantitative data; (b) making inferences and justifying conclusions; (c) conditional probability and the rules of probability; and (d) using probability to make decisions (p. 80). Each of these components has a heavy focus on statistical reasoning with no acknowledgment of the issues related to learning the necessary probabilistic concepts. Instead, probability is regarded as a set of rules to be learned (a sentiment evident in the title of the third component). Thus, overall the CCSSM appear to largely exclude probability and instead focus on statistics. The lack of emphasis on probability in the early grades is particularly troubling given that research has shown the formation of misconceptions about probability begin in students as young as 7 years of age (Fischbein 1975). Moreover, there is evidence that elementary students are receptive to instruction about probability topics. Such instruction could promote the development of secondary intuitions and curb misconceptions.

5 Final Thoughts

From our practitional perspective, we believe that the lack of emphasis on probability in the CCSSM shows a misalignment of research, learning theory, and practice. Analyzing the CCSSM shows a gap between the research findings and curriculum standards with regard to probabilistic thinking. It is puzzling that this has occurred given the current climate of accountability and research-based evidence influencing education in the U.S. (e.g., Feuer et al. 2002; Kelly 2008). Although our perspective may appear U.S.-centric, we note that researchers have reported a similar narrowing of the curriculum in England (Kapadia 2009). Moreover, a recent article by Schmidt and Houang (2012) indicated that the topic of probability is not included in the K-8 curriculum of two-thirds or more of the top achieving countries in the world (what they refer to as the A+ countries from TIMSS).

Our review of probabilistic frameworks from a practitional perspective has left us with several critical questions. First, if organized frameworks and models are not being used to inform curriculum standards, what chance do single studies on particular probabilistic ideas have of informing practice? Second, the history and development of these frameworks have been discussed, but what about the legacy of findings? The research base exists and there are existing frameworks that can be built on. Is there no one to champion the cause? Finally, the push for statistical thinking is prevalent in curriculum standards (e.g., CCSSM), yet one must ask, to what degree can one understand statistics without understanding probability?

References

Biggs, J., & Collis, K. (1982). *Evaluating the quality of learning: the SOLO taxonomy*. New York: Academic Press.

Biggs, J. B., & Collis, K. F. (1991). Multimodal learning and the quality of intelligent behavior. In H. A. H. Rowe (Ed.), *Intelligence: reconceptualisation and measurement* (pp. 57–66). Hillsdale: Erlbaum.

Borovcnik, M., & Peard, R. (1996). Probability. In A. Bishop, M. A. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education* (Vol. 1, pp. 239–287). Dordrecht: Kluwer.

Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly*, *104*(9), 801–823.

Feuer, M. J., Towne, L., & Shavelson, R. J. (2002). Reply to commentators on "scientific culture and educational research". *Educational Researcher*, *31*, 28–29.

Fischbein, E. (1975). *The intuitive sources of probabilistic thinking in children*. Dordrecht: Reidel.

Fischbein, E. (1987). *Intuition in science and mathematics*. Dordrecht: Reidel.

Fischbein, E. (1999). Psychology and mathematics education. *Mathematical Thinking and Learning*, *1*(1), 47–58.

Fischbein, E., & Gazit, A. (1984). Does the teaching of probability improve probabilistic intuitions? *Educational Studies in Mathematics*, *15*, 1–24.

Fischbein, E., Pampu, I., & Manzat, I. (1970). Effects of age and instruction on combinatory ability in children. *The British Journal of Experimental Psychology*, *40*, 261–270.

Greer, B. (2001). Understanding probabilistic thinking: the legacy of Efraim Fischbein. *Educational Studies in Mathematics*, *45*, 15–33.

Hawkins, A. S., & Kapadia, R. (1984). Children's conceptions of probability—a psychological and pedagogical review. *Educational Studies in Mathematics*, *15*, 349–377.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. London: Routledge & Kegan Paul.

Jones, G. A. (Ed.) (2005). *Exploring probability in school: challenges for teaching and learning*. New York: Springer.

Jones, G. A., Langrall, C. W., Thornton, C. A., & Mogill, A. T. (1997). A framework for assessing and nurturing young children's thinking in probability. *Educational Studies in Mathematics*, *32*, 101–125.

Jones, G. A., Langrall, C. W., Thornton, C. A., & Mogill, A. T. (1999). Students' probabilistic thinking in instruction. *Journal for Research in Mathematics Education*, *30*, 487–519.

Jones, G. A., Langrall, C. W., & Mooney, E. S. (2007). Teaching and learning probability: a research perspective. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 909–955). Reston: National Council of Teachers of Mathematics.

Kader, G. (1999). Means and MADS. *Mathematics Teaching in the Middle School*, *4*(6), 398–403.

Kahneman, D., & Tversky, A. (1972). Subjective probability: a judgment in representativeness. *Cognitive Psychology*, *3*, 430–454.

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

Kapadia, R. (2009). Chance encounters—20 years later fundamental ideas in teaching probability at school level. *International Electronic Journal of Mathematics Education*, *4*, 371–386.

Kelly, A. E. (2008). Reflections on the national mathematics advisory panel final report. *Educational Researcher*, *37*, 561–564.

National Governors Association for Best Practices, & Council of Chief State School Officers (2010). Common core state standards for mathematics. Retrieved from http://www.corestandards.org.

Piaget, J., & Inhelder, B. (1975). *The origin of the idea of chance in students*. New York: Norton. (L. Leake Jr., P. Burrell, & H. D. Fischbein (Trans). Original work published in 1951.)

Polaki, M. V., Lefoka, P. J., & Jones, G. A. (2000). Developing a cognitive framework for describing and predicting Basotho students' probabilistic thinking. *Boleswa Educational Research Journal*, *17*, 1–21.

Schmidt, W. H., & Houang, R. T. (2012). Curricular coherence and the common core state standards for mathematics. *Educational Researcher*, *41*, 294–308.

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.

Tarr, J. E., & Jones, G. A. (1997). A framework for assessing middle school students' thinking in conditional probability and independence. *Mathematics Education Research Journal*, *9*, 39–59.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, *185*, 1124–1131.

Volkova, T. (2003). *Assessing Russian children's thinking in probability*. Unpublished Master's thesis, Illinois State University, Normal, IL.

Watson, J. M., & Caney, A. (2005). Development of reasoning about random events. *Focus on Learning Problems in Mathematics*, *27*(4), 1–42.

Watson, J., & Moritz, J. (1998). Longitudinal development of chance measurement. *Mathematics Education Research Journal*, *10*, 103–127.

Watson, J. M., & Moritz, J. B. (2003). Fairness of dice: a longitudinal study of students' beliefs and strategies for making judgments. *Journal for Research in Mathematics Education*, *34*, 270–304.

Watson, J. M., Collis, K. F., & Moritz, J. B. (1997). The development of chance measurement. *Mathematics Education Research Journal*, *9*(1), 60–82.

Experimentation in Probability Teaching and Learning

Per Nilsson

Abstract This chapter explores the relationship between theoretical and empirical probability in experimentation-based teaching of probability. We examine previous research and a fresh small-scale teaching experiment in order to explore probability teaching, which involves students' (12–13 years old) experimentation with data.

The literature review and the teaching experiment point to several challenges for teaching probability through experimentations. Students emphasize absolute frequencies and part–part relationships, which makes it difficult for them to understand the principle of replacement and end up with numerical values to probability estimates. Students also find it hard to compare and make inferences if the samples are made up with different numbers of observations.

According to teaching strategies, the teaching experiment shows how experimentation encourages students to engage in questions of chance and probability. Among other things, it is also shown how variation of meaning-contexts supports students understanding of unfamiliar situations and how comparison-oriented questions can be used to promote students understanding of the relationship between theoretical and empirical probability.

1 Introduction

It is well known that both research and teaching in probability learning have had and largely still have their main focus on the theoretical, Laplace-oriented interpretation of probability (Jones et al. 2007). In recent years, however, this tradition has been questioned. Theoretical probability is based on the assumption of an equiprobable sample space and this is an assumption that can almost only be made in games of chance (Batanero et al. 2005). To act as conscious citizens of a modern society, we need to be able to express an opinion on the outcome of random events based on the observed frequencies of the event. However, there are also times when we must be able to express an opinion on the likelihood of an event when it is practically impossible to produce empirical evidence of the frequencies of a random event. We

P. Nilsson (✉)
Linnaeus University, Växjö, Sweden
e-mail: per.g.nilsson@lnu.se

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_28,

must be able to make predictions based on the composition of the random generator, although we cannot make assumptions about equiprobability. This dual nature of probability modeling supports a view of probability teaching that aims to develop an understanding of both theoretical and empirical (frequentist) probability and, in particular, the relationship between them (Nilsson 2009).

This chapter provides a discussion on teaching and learning probability in the relationship between theoretical and empirical probability. We examine a small-scale teaching experiment, which involves acts of experimentation. In particular, the purpose of the experiment is to explore and illustrate critical aspects of probability teaching, which involves students' concrete production of and experimentation with data. Developing a background for the teaching experiment, we proceed with reviewing previous research which highlights issues regarding the relationship between empirical and theoretical probability in situations of experimentation.

2 Previous Research

The theoretical interpretation of probability allows the calculation of probabilities before any trial is made. It implies the need for a sample space-oriented ratio-thinking (Hawkins and Kapadia 1984) where the probability of an event is obtained by the fraction of outcomes favorable for an event out of all cases possible (Borovcnik et al. 1991). The research tradition has made us aware of students' ability and inability to list all possible outcomes for an event and their understanding of how the underlying sample space regulates the probabilities of a random phenomenon (Chernoff and Zazkis 2011). The empirical probability is posterior in that the probability of an event is obtained from the observed relative frequency of that event in several trials (Borovcnik et al. 1991). Similar to theoretical probability, empirical probability also implies ratio-thinking, but this time in terms of determining the fraction of the number of times an event appears out of all trials made.

By reviewing research containing experimentation with data, I was able to distinguish two methodological and analytical directions regarding the bi-directional relationship between theoretical and empirical probability: from theoretical to empirical probability or from empirical to theoretical probability. Which direction becomes most prominent depends very much on whether the underlying sample space is known or hidden to the students. In cases where the theoretical model of a random generator is known from sample space considerations, students are challenged to reflect on how the model is *mapped* on the distribution of outcomes that the random generator produces (e.g., Pratt 2000; Horvath and Lehrer 1998; Nilsson 2009; Saldanha and Thompson 2002). In cases where the theoretical probability is hidden or impossible to determine from sample space considerations, research focuses on students' ability to make informal statistical *inference* about the empirical probability and how they connect frequency information to theoretical probability (e.g., Brousseau et al. 2002; Pratt et al. 2008; Makar and Rubin 2009; Polaki 2002).

Mapping direction: Theoretical probability → Experiment → Empirical probability

Inference direction: Experiment → Empirical probability → Theoretical probability

2.1 Mapping Direction

Being able to generate all the possible outcomes is essential for theoretically modeling the probability of outcomes for a random event. However, learning techniques to generate all possible outcomes in a sample space has little value unless students realize the importance of how the sample space regulates the likelihood of random events and is reflected in data (Jones et al. 2007). Horvath and Lehrer (1998) show how children in grade 2 (ages 7–8) and grades 4 and 5 (ages 9–11) are able to develop an understanding of sample space mapping on data-distribution after being introduced to a notational system for organizing the sample space. When the experimentation differed from their predictions, the grade 2 children turned to the data instead of sample space considerations for making new predictions. That is, in a natural manner, they changed to the inference-direction. The older students, on the other hand, immediately incorporated the importance of the outcome space for predicting how the data would be distributed. The study stresses the importance of teaching students to systematically structure the sample space to enhance their capacity to understand and predict empirical distributions.

Pratt (2000) reports how students elaborated their understanding of the total of two dice in a *Chance-maker microworld* where they were offered the opportunity to generate and act on empirical evidence. Pratt's results showed that, as the students (ages 10–11) took notice of the frequencies of outcomes and modified a gadget's working box to influence the appearance of its associated graphical display, two new meanings (global resources) for the behavior of the totals of two dice emerged, the *distribution resource* and the *large number resource*. The distribution resource connects strongly with the 'mapping-direction,' similar to the methodology of Horvath and Lehrer (1998). It concerns the relationship between theoretical probability and frequencies in terms of how students develop their understanding of the role of the underlying sample space when trying to regulate the behavior of frequencies. Expressed in features of the Chance-maker microworld, Pratt and Noss (2002) describe the distribution resource as "the more frequent an outcome in the working box, the larger its sector in the pie chart" (p. 472). The large number resource builds on an understanding of the principle that (relative) frequencies stabilize as a function of an increasing amount of data. In the setting of the Chance-maker microworld, the resource was expressed as "the larger the number of trials, the more even the pie chart" (Pratt and Noss 2002, p. 471). Aspinwall and Tarr (2001) and Stohl and Tarr (2002) support Pratt's findings, also showing how children develop their understanding of the role of the number of trials in the relationship between theoretical and empirical probabilities when they have the chance to run as many trials as they want and view the results in graphical representations.

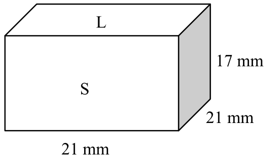


Fig. 1 Index *S* refers to the four small faces of the dice and index *L* to the two large faces

The dice were numbered: $D_1=\{2_L, 2_L, 2_S, 4_S, 4_S, 4_S\}$ and $D_2=\{3_L, 3_L, 3_S, 5_S, 5_S, 5_S\}$.

In Nilsson (2009), eight students (ages 12–13) in four teams played a dice game based on the sum of two dice. The teams were instructed to distribute a set of markers numbered from 2 to 12 on a playing-board. If one or both teams had at least one marker in the area marked with the sum of the dice, they removed exactly one marker from this area irrespective of which team rolled the dice. The team that was the first to remove all its markers from the board won. The game consisted of a system with four sets of dice, with an asymmetric shape and numbered in specific ways (Fig. 1).

During the activity, all teams progressed from a uniform distribution of the markers to a non-uniform distribution. Two of the teams based this progression on a developed understanding of how the characteristics of the random generator will be mapped in the series of throws in a game. In contrast to the activities of Pratt (2000) and Stohl and Tarr (2002), the dice-game is analog. In this way, it was not as easy to let the students generate and reflect on large series of outcomes. However, through systemic variation (Marton et al. 2004) and by designing random generators that generate clear differences in data distribution, the students were, nevertheless, challenged to develop this progression. In particular, it became very clear to the students that a total of nine, the largest total of round one, was difficult to get in comparison to the other totals. The conflict between the uniform distribution that the students were expecting and the non-uniform distribution obtained supported the students in their further understanding of the game.

2.2 Inference Direction

Research that falls within the frame of the inference direction is not as extensive or developed as the research reported on above. This is also one of the main reasons for conducting and investigating a fresh teaching experiment on this issue in the present chapter.

When we talk about the inference direction, we refer to activities where the underlying mechanisms of the random generator are either impossible to identify or are hidden. In the present chapter, the focus is on learning regarding the bi-directional relationship between Laplace-based probability and empirical probability. For this reason, we limit ourselves to a situation where we are able to model the underlying random process by sample space and geometrical considerations of the random generator. The purpose is that students should learn how to make probability generalizations that go beyond the data collected (Pratt et al. 2008), how to make inference from data to the features of the random generator producing the data, and the theoretical probability distribution the random generator implies (Makar and Rubin 2009). Therefore, we will look at a teaching situation that allows the students and the teacher to compare the generalizations to theoretical probability. It is important to note, however, that theoretical probability is not the 'true' probability. Nonetheless, it is reasonable to assume that, when using dice, coins, and urns, the theoretical model of probability gets close to the true probability (cf. Konold et al. 2011). In addition, our focus is on situations described by Pratt et al. (2008) "where the population cannot be described in terms of a total finite dataset but is most adequately described through a probability distribution" (p. 108).

Makar and Rubin (2009) report that frequency data are often disregarded by students (ages 7–9). Regardless of sample size, they found that "students struggled to connect conclusions to the data they have collected" (Makar and Rubin 2009, p. 95). Pratt et al. (2008) challenged students aged 10 to 11 to infer from data the unknown configuration of a virtual die and, their study also highlights students' difficulties interpreting information of a data sample (see also, Stohl and Tarr 2002). However, supported by the situation, the students in the study by Pratt et al. started to reflect on the data. What became critical for the students was to understand the difference between drawing conclusions based on the global (long-term) and local (short-term) behavior of a data sample.

Conducting two teaching experiments, including children of ages 8 to 10, Polaki (2002) shows that it is not always sufficient to give students the opportunity to generate a large amount of data (long-term behavior) to support them in making inferences about the underlying sample space from data. In one group, children acted on a small set of data while in another group the children were called on to reflect on sample space composition in relation to a large set of empirical data. Both instructional methods had a positive effect on the students' ability to link sample space composition to the probability of an event. Polaki (2002) conjectured that this might depend on the *limit process*, implied in the law of large numbers, being too abstract for the students and the fact that they were not familiar with computer-generated data.

Urn-like situations are often used to provide students with an unknown sample space. In Stohl and Tarr's (2002) study, sixth grade students were challenged to use simulations and data analysis to infer information about the content of a bag of marbles. The students knew how many marbles there were in the bag (known population size), but did not know how many marbles there were of each kind. Similar to the findings of Brousseau et al. (2002), the information about the number of marbles in the bag affected the number of trials conducted. In both studies, several students drew samples the same size as the number of marbles in the bag. Based on their findings, Stohl and Tarr (2002) developed a situation with an indeterminate population size where the students were asked to make an inference about the probability of getting a fish in a lake rather than merely counting each type of fish. Similar to the fishing situation, the current teaching experiment will include a task in which the sample space is hidden and the population size is indeterminate.

3 Experimentation in a Teaching Experiment

The purpose of the teaching experiment is not to evaluate successful teaching. Of course, there is an intention to stimulate, rather than discourage, the learning of probability. However, the primary purpose of the experiment is to explore and illustrate critical aspects of probability teaching, which involve students' concrete production of and experimentation with data.

3.1 Analytical Approach—Contextualization and Variation

The present teaching experiment focuses on students' understanding of concepts and principles related to the theory of probability. The analytical focus will be close to the mathematics involved in the teaching and, in particular, the students' meaning-making of the mathematics. We will draw on the theoretical idea of *contextualization* (Halldén 1999) and the theory of variation (Marton et al. 2004) to account for the design and re-design of activities in the teaching experiment and the meaning-making that appears during the activities.

The framework of contextualization constitutes a constructivist approach for taking into account the situational character of cognition (Ryve 2006). To speak about students' processes of contextualization is to speak about how learners struggle to render a phenomenon or concept intelligible and plausible in personal (cognitive) contexts of interpretation (Halldén 1999; Nilsson 2009). This idea rests on the principle that we always experience something in a certain way, from a certain set of premises and assumptions (Säljö et al. 2003). For instance, we can understand a car in different ways depending on from which context we consider it. We may consider it as an object of conveyance. However, we can also talk about it as an element in the development of our industrial society, or as how it might work as a status symbol in a social context. Which aspects emerge, how we experience and develop an understanding of a phenomenon is based on the context from which we choose to consider the phenomenon. It is important to stress, therefore, that context here does not refer to the spatiotemporal setting of the learning activity but to a mental device shaped by personal interpretations of the activity.

The main reason why we find contextualization particularly useful as a tool for analyzing the teaching and learning situations of probability is that probability itself implies a variety of possible contextualizations. First, the theoretical and empirical interpretations of probability constitute two general mathematical and formal contexts for the interpretation of probability. Second, many of our notions of chance and probability are developed from our everyday-experiences which teachers can use for discussing and making sense of ideas belonging to the more formal contexts.

If we assume that understanding develops from a process of contextualization, teaching needs to offer students the possibility to develop contextualizations in which a mathematical treatment appears to be relevant and meaningful to them. Contextualization also implies that, in cases where students find the mathematics difficult and incomprehensible, it may help if the student is challenged to consider it from another context. We talk of *variations of translation* when talking about situations where the context is changed in a teaching activity in order to elaborate on a mathematical idea. The mapping direction and the inference direction discussed earlier constitute two forms of contextual variations. Another principle of variation that will be used as an analytical means is *variations of discernment*.

Variations of discernment occur when variation becomes a context for interpretations per se. In contrast to variations of translation, where a mathematical idea is transferred between contexts, e.g., between a mathematical context and an everyday context, variations of discernment keep to the same context and, in that context, a mathematical idea is made sense of by varying certain aspects of the idea while keeping other aspects constant. Take a concrete example. The dice-game used in Nilsson (2009) constituted a setting in which the students were able to experience the meaning and relevance of chance and probability. In the study, it also became apparent how the students associated the game with their everyday experiences of board games and how they changed between a theoretically oriented context and a frequency-oriented context to increase their chances of winning the game. In these senses, they expressed variations of translation. However, the game was repeated four times during the activity and each time the students played with a new pair of dice. The systematic variation in the design of the dice between the rounds supported variations of discernment, enabling the students to discern the critical aspect of favorable outcomes for the probability of the totals.

The idea of variations of discernment is heavily influenced by the theory of variation (Marton et al. 2004). Another idea from the theory of variation that we find fruitful for our investigations is the notion of the *enacted object of learning*. The enacted object of learning is an interactional phenomenon and part of the classroom interaction (Ryve et al. 2011). To analyze the enacted object of learning includes revealing the opportunities and limitations for a learner to experience features of the learning object as it is established in the classroom interaction (Runesson 2005). Hence, in conjunction with the analysis of the teaching experiment, we are interested in what contextualizations and patterns of variation emerge in the classroom interaction. This is not to say that all students actually reach an understanding of the content involved in the interaction. What we are expressing an opinion on is the enacted object of learning, that is, what is possible for students to learn in relation to their contextualizations and the patterns of variation of the content involved in their contextualizations.

3.2 Data Collection

The teaching experiment was conducted in a grade 6 class (ages 12–13) at a Swedish primary school. The students are familiar with traditional Swedish teaching of mathematics with elements of group-work and whole-class discussions. However, experimentation is unusual; the textbook provides the basis for the teaching of mathematics. The students have not been formally introduced to probability in school. They are familiar with decimal numbers and fractions and have some experience of percentages.

The teaching experiment consisted of five sessions. In the first four sessions, the class was split in two halves (Group A and Group B) with ten students in each group. Within these groups, the students were organized in five small groups with two students in each group. The author of the chapter acted as the teacher during the entire experiment and is referred to as the investigator (I) throughout the rest of the text. The teaching switched between group-work and whole-class discussions. Using the interactive whiteboard, diagrams were prepared in advance. In these diagrams, the results of the experimentation from all five groups could be presented to the whole class, serving as a platform for communication and for discerning patterns of variation.

Because we could only arrange the use of four video cameras, one of the groups was only audio-taped. One of the group-cameras was also used for documenting whole-class discussions. A research colleague assisted with the practical work of switching between group discussions and occasions when the teacher led whole-class discussions.

3.3 Teaching Plan

In the following, we will focus on the four first lessons. The last, fifth lesson, served mainly the role of summing up and concluding the entire teaching experiment. The four first lessons were conducted in the following order:

Lesson 1A, Lesson 1B, Lesson 2A, and Lesson 2B.

The letters A and B refer to the groups of students involved in the lessons. Numbers 1 and 2 illustrate that the entire teaching experiment consisted of two overall themes, aiming to challenge students to explore different aspects of the bi-directional relationship between theoretical and empirical probability. Lesson 1B was adjusted, based on the analysis of Lesson 1A. The experiences gathered from Lesson 1A and Lesson 1B then served as a basis for adjusting the planned Lesson 2A. By analyzing Lesson 2A, Lesson 2B was further specified.

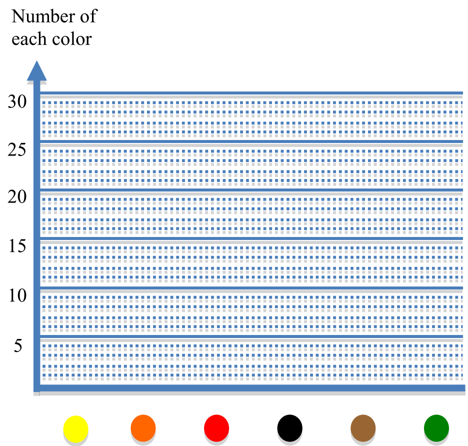
Before starting the analysis, we present the overall tasks of the lessons. Details of the lessons appear in the presentation of the results and analysis of the lessons.

3.3.1 Lesson 1

The activity that formed the basis for the teaching experiment was to determine, through experimentation, the probability of picking a piece of candy of a certain color from a bag of candy. In Sweden, the candy is called 'Non Stop'.[1] There are six different colors of candy in the bag, all pieces are the same shape, and the brand is well known to the students.

Central to the teaching was that the students would produce their own concrete data. In this way, their reflections could be grounded in data that would have a

[1] 'Non Stop' candy is similar to 'Smarties' in the UK or 'M&M's' in the USA.



Color	Quantity	Probability (percentage)
Yellow		
Orange		
Red		
Black		
Brown		
Green		
Total		

Fig. 2 Diagram and table for the recording of observations

history for them (Cobb 1999). The five groups each received one bag of candy in Lesson 1 and were told by the investigator that he would very much like to have a piece of yellow candy. The driving question (Makar and Rubin 2009) was to help the investigator to select from which bag the candy should be picked in order to increase the chances of obtaining a yellow piece. The students were told that they were only allowed to pick one piece at a time and had to put the piece back in the bag before they picked another piece. Hence, the situation included replacement which would allow for creating an infinite population or process in order to get at the underlying probability distribution. The bags were of equal size (according to their weights) but the exact number of pieces is not determined (cf. Stohl and Tarr 2002) and would probably be different between the bags.

The students were first asked to record their observations in a diagram and then asked to record their data in a table (Fig. 2).

The overall intention of the lesson was to develop an understanding of sampling and relative frequencies. In the next lesson, the students were supposed to compare their empirical estimates of probability to the actual contents of the bag and the theoretical probability distribution. Relative frequencies, percentages, and proportions were principles that were intended to provide contexts for such comparisons.

3.3.2 Lesson 2

The first activity in Lesson 2 followed on from Lesson 1. The students were supposed to reflect further on the difference between absolute frequencies and relative frequencies. In the next phase of the lesson, the students counted the pieces of candy in their bags. They were asked to reflect on how the percentages of the different colors in the bag correlated to the percentages obtained in the sample.

The lesson ended with all groups presenting their table results (Fig. 2) on the interactive whiteboard, for use in a discussion on the relationship between theoretical and empirical probability.

4 Results and Analysis

4.1 Lesson 1A

When introducing the lesson, the investigator tells the students that he will act as the teacher for the class during the teaching experiment lessons. Their regular teacher is only in the classroom as an observer.

4.1.1 Sample Space Oriented Contextualization (Quantity Counting)

One bag of candy is distributed to each of the five groups and the students are asked if they can tell which bag should be chosen if we want a yellow piece and if we are only allowed to pick one piece from one bag without looking. The investigator continues by asking if the students think it matters which bag is chosen. Sergio answers that "Maybe it [the choice of bag] contains… it maybe contains more yellow." The question whether it matters which bag is chosen implies variations of discernment between the bags. Supported by this variation, Sergio contextualizes the chance of drawing a yellow piece in a sample space context. When the class is asked to reflect on what Sergio is saying, they reach an agreement that it is important to take into consideration the quantity of yellow pieces in a bag when deciding which bag to choose.

The investigator concludes that we now have an idea, that it is important to know which bag contains most yellow pieces. "The problem is, though," he says, "we do not know what the contents of our different bags look like and we are not allowed to empty out the bags and count the candy."

The class is told that they should only draw one piece of candy at a time and then put it back in the bag before drawing again and making a new observation. They are also told that they are allowed to repeat this procedure as many times they like. The students are encouraged to discuss in their groups how to do this. None of the groups raises questions about sampling, data, or proportions. The proposals made by students are all in sample-space oriented context, centering on the idea of counting the quantity of the different colors in the bag. Carl, for instance, suggests that we can assemble them in piles according to how we pick them. What later becomes evident is that the piles he is thinking of are the concrete piles of all of the candy. When he is done with his work, the bag is empty. The idea does not refer to any kind of symbolic notation of observations. The groups are reminded that they should replace the candy after each observation. This is, however, a situation that is hard for the students to understand. The task the students see as their obligation to solve is to figure out how many pieces of candy there are in the bags and not the proportions of candies in the bag or how often a color will appear in a sample with replacement. Monica and Lars express the problem their sample space oriented 'counting' contextualization causes them in trying to make sense of the situation; the investigator describes:

Monica: Isn't there a risk that you pick the same one again?
Lars: Yes, exactly!

Even though the investigator notices that most of the students find it hard to understand the idea of the experiment, he introduces the experimentation activity: "We are starting to do [fill in] these diagrams (Fig. 2) and we'll see... and so we can think a bit more when we have it in front of us, what we have been getting out." One reason that the investigator lets the students start their experimentations in their groups even if the students do not understand the situation and its purpose completely is because he wants to see how the experimentation per se could encourage the students to ask questions about chance and encourage them to reflect on how the data of the experiment could be used for understanding the chances of the different colors (cf. Shaughnessy 2003).

The groups start to draw pieces of candy and enter their observations in the diagrams. The discussion is limited. None of the groups reflect on issues related to empirical probability, such as, reflecting on part–part variations in the absolute frequencies between the colors or on part–whole variations in relation to how many trials were done. Instead, several students make comments that disclose that their way of experiencing the situation is still guided by a contextualization which is about counting how many pieces there are of each color. Another aspect of this issue is that most of the students mix the bag of candies more carefully after having observed the same color in a row. Among other things, we could consider this also to be an expression of their wish to minimize the risk of drawing the exact same piece of candy more than once, since this would be unfair to the other colors, regarding the idea of merely counting the quantity of pieces of each color in the bag.

4.1.2 The Problem of Part–Whole Reasoning

When all groups are done with their experimentations, they are asked to reflect on what the chance of getting a piece of yellow candy would be between the bags. In particular, they are asked whether they can come up with any numerical values that could be used for comparison. None of the students respond and the investigator interprets the situation as though the students are unable to conceptualize the meaning of chance. Reminded of the students' natural understanding of the role of sample space composition for the chance of an event, he chooses to frame the discussion of chance based on an every-day context regarding the probability that a butterfly flying in through the classroom window lands on a girl:

I: How big is the chance that he lands on a girl?

Lars responds:

Lars: In that case, it is how many girls there are against us three [boys].

The situation is clear to the students. The chance is equal for all individuals in the classroom. None of the students raise any subjective issues such as it would depend on the smell of a person, the color of one's clothes or how close to the window one is sitting. Instead, and similar to Lars, several students seem to make sense of the situation by noticing a part–part oriented context; the part of boys in comparison to the part of girls. None of the students refer to aspects of part–whole reasoning. However, the situation offers the investigator a context that makes it possible for him to give meaning to percentages and to link percentages and chance:

I: What is the percentage of boys here? How big is the chance that it [the butterfly] lands on a boy? It is… [humming] percentage chance. How many percent is that, how can we calculate that?

In the discussion, students express an understanding of what a part is and that boys constitute one part and girls constitute one part. Still, they have difficulty relating the parts to a whole in a numerical context. In previous teaching on percentages and fractions, they have worked with images and material such as cakes, pizzas, and chocolate bars, talking in terms of *part of a continuous quantity*. The present case is about the *part of a discrete quantity* (Charalambous and Pitta-Pantazi 2007). To give an idea of this, the investigator continues:

I: What is 100 % of us here [Moving his arms in circles over the class]?

After some discussion, the class concludes that 100% is all of us and that we are 13 people (four boys and nine girls). The investigator writes on the whiteboard that we are 13 in total in the classroom and then he asks what percentage is boys.

Monica: If we think that we are 13, then one can think that each person is one part. Then, that is it, four parts are boys and nine parts are girls.

Even if Monica's reasoning contains fraction thinking, it is still unclear whether she is thinking in terms of part–part comparison or in terms of part–whole comparison. The investigator asks Monica to continue her line of reasoning and he helps her formulate that the boys are four parts of 13. This is not done just for Monica's benefit, but also to make the calculation of the percentage an enacted object of learning for the entire class.

The students use calculators to conclude that there are 30% boys in the classroom. The investigator then takes this opportunity and translates this to a chance context concluding that:

I: A butterfly, thus, has a 30 % chance of landing on a boy.

Adopting the investigator's language, Carl concludes that there is a 70 % chance that the butterfly lands on a girl. Under guidance, the students translate the reasoning and calculations done in the butterfly context to the context of their observations. The groups calculate the relative frequencies of each color in their sample and their results are recorded on the whiteboard. Thus, the whiteboard displays all the five groups' diagrams. Based on what the whiteboard displays, the students are asked which bag to draw from. Carl replies "I say Lars and Sibel's, given that they have

25 % and it is... more than the others. One [group] has nine, one has 14, one 25 [Lars and Sibel's], one has 19, one has 15 and then there is 25, which is most."

The preceding discussion, developed in accordance with the butterfly context has, of course, played a major role in getting the students to use percentages as a means for comparing the bags. Before the butterfly contextualization, none of the students put forward any proposals for a numerical comparison. What happened can be considered as a loop of variations of translation. First, the question of how we can compare chances between the bags of the groups was translated to a butterfly context. In this contextualization, they were able to make sense of percentages as a way of measuring chance. This understanding was then translated and used to make sense of the groups' observations in the candy situation. From Carl's comment, we can also see how the variation displayed on the board helps him to discern and give meaning to percentages as a measure of chance. As such, the displayed variation provided him with material for making inferences on the content of the bags.

4.1.3 Fairness Thinking

Monica, however, is not completely convinced. She thinks that it is impossible to make a choice since all of the groups have made different numbers of observations. She argues that one group which drew candies few times would have perhaps observed more yellow if it had continued drawing. The reasoning implies part–part reasoning. It is focused on the part of yellow without any explicit considerations that the total also increases if the sampling is increased. However, the reasoning also implies a thinking of fairness. Following up on Monica's comment, Carl says:

Carl: It should be as equal [fair] as possible, then everybody actually would have made the same number of draws. Then it is easiest to judge.

The investigator challenges this by asking the students if it would have mattered if *all* groups developed a sample of six observations or 139 observations. The question was also an attempt to introduce the theme of the next lesson on the role of sample size. This extreme variation, however, does not help Carl to discern any difference:

Carl: No, but really, if everyone draws the same then all have the same to compare with.

Pratt (2000) speaks about fairness as a resource for judging whether a random generator produces outcomes with an equal chance. Similar to Pratt's conceptualization of fairness, Carl's reasoning implies issues of equality. Nevertheless, this time it is not about equal chance but about equal conditions for making a fair judgment based on the data. However, such an understanding of fairness is in conflict with the formal idea of the law of large numbers and confidence intervals. Comparing two samples with only six observations each is less fair than comparing a sample of 100 observations with a sample of 200 observations. We can express an opinion with a higher degree of certainty in the latter case.

4.2 Lesson 1B

When analyzing the first lesson, it appeared that the students did not, in a more or less spontaneous way, understand what the experiment was about and what the frequency information obtained from the experiment would provide them with. On account of that, the investigator starts Lesson 1B by introducing a discussion on the part–whole relationship in the underlying sample space and the meaning of a sample, what a sample can be used for.

4.2.1 Stressing the Notion of Chance

Similar to Lesson 1A, the class turns into a discussion of the role of sample space, i.e., of the content of the bags. However, this time the investigator frames his contributions to the discussion more often in terms of 'chance'. He holds up a couple of bags to illustrate that they are of equal size. Then he asks the students to consider the difference (variation) in the chance of getting a yellow piece from a bag with ten yellow pieces and a bag with 100 yellow pieces. Melissa responds, "You have a higher chance of getting yellow if there are more yellow in the bag." Based on Melissa's suggestion, the investigator writes on the whiteboard "Higher chance of getting a yellow in the one (bag) with the most yellow." This definition implies part–part thinking. To get the students to reflect on the situation from a part–whole oriented context, they are challenged to reflect on which bag they would choose if they knew that in a bag with 200 pieces there are 25 yellow and in a bag with 250 pieces there are 26 yellow. None of the students suggest the bag with 250 pieces. Katie represents the discussion when she concludes, "If you take the one with 200 pieces, there are 25 parts in it. But, in the one with 250 pieces there are 50 pieces more, but there is only one more yellow than the one with 200 pieces." The variation used in the question helps the students to discern and formulate arguments on aspects of probability in terms of proportions and not only in terms of absolute values. Based on the discussion, the students are asked to use percentages to reformulate the definition on the whiteboard in order to draw a conclusion from what we have come up with. Roger articulates: "There is a higher chance of getting a yellow where there is the highest percentage of yellow."

In Lesson 1A, the notion of a sample was not clear to the students. Therefore, to illustrate the meaning of a sample, Lesson 1B begins by considering how we can come to an understanding of the chance that a person in Sweden owns a mobile phone. The students understand that we cannot ask everyone in Sweden. Nevertheless, they show a natural understanding that we can base our decision on a sample of the population of Sweden. The mobile phone example concerns a finite population and the candy situation concerns an infinite population since it is a situation with replacement (Pratt et al. 2008). However, neither the investigator nor any student reflects on this difference between the situations as the mobile phone example is translated and used as an introduction to the experimentation with the bags of candy.

4.2.2 Talking About Percentages and Chance Through the Variations of the Samples

The groups perform their experimentations and fill in their observations in the diagram and the table (Fig. 1). Then the investigator asks the students to record their results in the diagrams prepared in advance on the board. Hence, similar to Lesson 1A, the class is able to see the results of all groups' experimentations.

In the picture on the whiteboard, one can discern that Roger's yellow bar is proportionally greater than the yellow bars of the other groups. This may also be how the students are thinking. However, it is not explicit in the way they express their argument. They say they chose Roger and Elisabeth's bar because it is the largest. However, it is not clear if they mean largest in an absolute or in a relative meaning. Focusing on this issue, the investigator asks the class if they can come up with the value of the chance of picking a yellow piece from each group respectively, based on the samples. Before the investigator once again makes use of the butterfly example in this lesson, he stresses the concept of chance further by asking the students how big the chance would be of getting a yellow if the bags only contain yellow, contain no yellow, or contain half yellow. No one in the class has difficulties in articulating that the chances would be 100 %, 0 %, and 50 %, respectively. Based on this, the investigator turns to the butterfly example to make sense of how we can calculate percentages as a quotient between a part and the whole (part of a discrete quantity).

When all of the groups are done with their calculations, they are asked to consider the questions: "Who has the highest percentage? Who got the highest number of yellows expressed as a percentage?" To emphasize the meaning of the sample, the investigator continues before the students have had a chance to answer, "Who most often got yellow in their sample? That is, what percentage of the time did you get a yellow?" These quantities are not recorded on the board. Each group orally presents the percentage of yellow obtained in their sample. There is no doubt that they would choose Roger and Elisabeth's bag since they had observed yellow 23 % of the time, which was the highest value. The lesson ends with questions regarding the students' decisions regarding other colors. Without any reflection on sample size, the class establishes the practice (cf. Cobb 1999) of choosing the sample that shows the highest percentage of the color in question.

The fairness thinking that appeared at the end of Lesson 1A regarding sample size never became an issue in Lesson 1B. However, a kind of fairness thinking did appear, but this time in regard to the draw procedure. In Lesson 1A, several students saw a risk of picking exactly the same piece several times. In Lesson 1B, one of the students, Melissa, mentioned this issue but got no support from the other students. Actually, Roger, using a variation of discernment, explained to her that "That could be the case for me as well," after that Melissa said that she might have picked the same piece several times. In this case, the design of the activity provided Roger with the opportunity to refer to variations between the samples. Another variation of discernment to apply when making sense of the risk of drawing the same piece of a target variable could be to ask whether this risk is as large as it is for all variables (colors) in an infinite population/process.

4.3 *Lesson 2A*

4.3.1 Discerning Relative Frequencies

When analyzing Lesson 1B, the positive effect of letting the students reflect on the difference in chance of picking a yellow candy from a bag of 200 pieces with 25 yellow or from a bag of 250 pieces with 26 yellow was recognized. The effect was positive in the sense that it turned the students' attention to part–whole relationships. Therefore, at the beginning of Lesson 2A, a similar task was used to strengthen the discussion about proportions and percentages. The groups' diagrams were rewritten on a computer and now projected onto the interactive whiteboard (Appendix A). All groups also received a printout of the computer-picture. On the printout, the students were asked the two following questions:

1. Suppose that you would like a piece of brown candy. You are only allowed to pick one piece without looking from Sara and Sergio's bag or from Carl and Louise's bag. Which of the bags would you choose? Explain you answer.
2. You would like a piece of red candy and you can choose which bag you want it from. Which of the groups' bags would you choose? What is the probability of getting a red piece from the choice you made?

The colors and the groups are not selected randomly. Sara and Sergio reached a larger absolute value in comparison to Carl and Louise. But the relative frequency of Carl and Louise's value is larger. Most students realize that a group may have got more of one color because they repeated the sample more times. They realize that they cannot only look at how many times a color was observed. They have to consider how many times it appeared in relation to the total number of observations. As Lars and Sibel discuss in their group:

Lars: There is a higher chance for us to get it [a brown] here [pointing to Carl and Louise's diagram].
Sibel: Yes.
Lars: They only had 81!
Sibel: Yes.
Lars: But, they [Sara and Sergio] drew more [brown], but that was because they had almost 50 more observations.

Similar to Lars and Sibel, and supported by the variations implied in the two tasks, several students started to discern a contextualization that emphasizes part–whole relationships. In the first of the two tasks, Lars seems clear that he will make his choice in accordance with the highest percentage of the target color. However, his understanding seems to be weak. Turning to the second task, he argues that they should go for Lisa and Eva's bag since they have the most observations of red. Initially, both Lars and Sibel note that two of the groups have observed red 21 % of the time. They also note that one group (Lisa and Eva) obtained more observations of red but that they had made 29 more observations than Linda and Jill. However, for some reason, that does not become clear from the group's discussion, Lars convinces Jill to go for Lisa and Eva's bag.

All of the groups, except for Lars and Sibel, have based their choices for both tasks on a comparison of the relative frequencies. When they are asked to explain their choice for the second task, Lars also changes his choice to Monica and Jill. There is no discussion as to whether it is better to choose Monica and Jill's bag or Carl and Louise's bag. Both groups observed red 21 % of the time. However, the episode shows how teachers can create opportunities for students to discern the meaning of samples and relative frequencies by systematic variations based on the students' history of experimentations with random phenomenon.

4.3.2 Coordinating Relative Frequencies to Sample Space Proportions

In the next task, the students were asked to calculate the actual (theoretical) percentages of the colors in the bag and compare them with the percentages they received in their samples. The goal was to call for reflection on the relationship between theoretical and empirical probability and the role of the sample size in this relationship. After completing their calculations, the groups recorded their results on the whiteboard, displaying all of the groups' tables (Fig. 2) in the same picture.

Unfortunately, the results of the experiment did not offer the possibilities that were expected, i.e., to create a context for discussing the role of sample size in the relationship between theoretical and empirical probability. However, for the present analysis this is not a problem as the primary purpose of the teaching experiment is not to evaluate successful teaching. The purpose is to reach an understanding of the critical aspects of experimentation-based teaching of probability. Hence, we claim that one such critical aspect is the creation of distinct differences in sample size, in order to create a context for discussing the role of sample size for the bi-directional relationship between the proportions of a sample space and the relative frequencies in a sample.

The experimentation in Group B was already done and the samples they produced did not differ in sample size any more than the samples produced in Group A. For that reason, it was decided to tone down the discussion on the role of sample size between the groups. Instead, the investigator chose to restrict the discussion to look at the relationship between theoretical and empirical probability within each group's sample. In practice, this was dealt with by trying to establish a pattern of variation in which the three most frequently observed colors were compared to the three largest proportions of colors in the bag. The possibility of this pattern of variation of discernment was accounted for before Lesson 2B in that the investigator had counted the content of the group's bags and found this possible pattern to be discernible.

4.4 Lesson 2B

The lesson begins in the same way as Lesson 2A did. The students reflect on which bag they would choose given a certain color;[2] reflecting on the importance of rel-

[2]Of course, this time the tasks were framed in terms of the samples conducted in Group B.

ative frequencies rather than absolute frequencies which were discussed during the group's first lesson. Now, this is reinforced, and all groups base their conclusions on differences in percentages.

To strengthen the focus on correlation between empirical and theoretical probability from an inference direction, attention is drawn to Katie and Allan's sample (Appendix B). The class is asked which color they think there is most of in Katie and Allan's bag. Except for one group, Peter and Charlie, the class hypothesis is that there are mostly green and orange in their bag. When Charlie and Peter are challenged to explain their answer, it appears that they only came up with something without any really consideration.

The students calculate the contents of their bags and the percentage of the different colors. The class is gathered and each group records their results on the whiteboard, giving the picture of five recordings in Fig. 2. In contrast to Lesson 2A, this time the investigator tried to restrict the discussion to the results of one group at a time. Taking their hypothesis as a point of departure, a discussion developed comparing the percentages of the sample of a group with the percentages of the colors in the bag.

By distinguishing between the three most frequent colors and the three least frequent colors, the investigator was offered a pattern of variation through which he could direct students' attention towards the relationship between the two probability contexts. During this discussion, the students strengthened their hypothesis regarding the content of Katie and Allan's bag.

As the students now have a sense that the proportion of a color in a bag correlates to the proportion of the color in a sample, the investigator calls their attention to the fact that the percentages do not coincide exactly. The question he poses to the students is if they can come up with any suggestions as to how we can make the empirical probability approach the theoretical probability further. The investigator makes it clear to students that this can be achieved. The question is how.

I: Talk to each other, discuss if you should take fewer [smaller samples] or if you should take more or how you are going to do it.

It is not a simple question for the students and, as was the case in Lesson 2A, none of the groups spontaneously suggest that they will get closer if the sample size is increased. Moreover, the experimentation offers no distinct pattern of variation which could be used to guide the discussion towards the critical aspect of sample size. Roger proposes the idea that they should develop small samples and when the two rates of percentage are close they should stop. This might not be a bad idea, depending on how you contextualize the question. Consider, for example, the situation of tossing a coin. If you get one head and one tail, you stop. The theoretical and empirical probabilities are perfectly matched!

5 Discussion

This chapter has provided a discussion on experimentation in the teaching and learning of probability. Based on a review of previous research, a small-scale teaching experiment served as a context for exploring and illustrating critical aspects for teaching probability, using the students' own production of and experimentation with concrete empirical data.

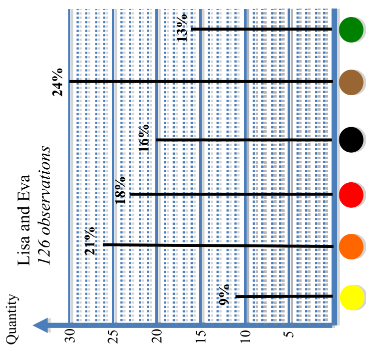
Cobb (1999) emphasizes the importance for students of acting on data that has a history for them, such as being grounded in the situation and reflecting particular purposes and interests to the students. This is consistent with Shaughnessy's (2003) call for introducing probability through data, which is a call the present teaching experiment gives further support to. The students' self-produced data and the variations between the samples provided contexts through which the teacher and the students could illustrate and develop a discussion of concepts and principles related to theory of probability. However, what becomes apparent during different instances of the teaching experiment is that experimentation-based teaching of probability is not an easy enterprise and that experimentation in itself does not necessarily encourage the students to reflect on the purpose of data, see statistical information as useful evidence for getting a picture of a population, or make probability predictions for random situations (cf. Makar and Rubin 2009). As such, the chapter in general and the current teaching experiment in particular point to the need for further investigations of probability learning and teaching strategies in experimentation-based probability teaching.

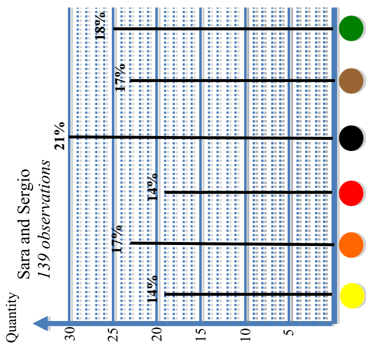
Brousseau et al. (2002) raise some issues as to whether it really is a good idea to explicitly compare empirical probability and theoretical probability in the teaching of probability. The authors claim that the comparison runs the risk of 'determinationalizing' empirical probability if it always is checked against the reality (theoretical probability). The present teaching experiment emphasizes that we have to be alert to the dominant role a theoretically-oriented contextualization may take in the students' meaning-making in an experimentation activity which aims at developing students' understanding of the bi-directional relationship between empirical and theoretical probability. For instance, at the beginning of Lesson 1A, the enacted object of learning concentrates on whether it matters from which bag one draws if one wants a yellow piece of candy. Because of this discussion, several students focused on counting the contents of the bag (cf. Stohl and Tarr 2002). Frequency information did not appear as a context for reflection on its own. Some students did actually consider the absolute frequencies of a certain color in the sample as the actual number of pieces of the color in the bag. In conjunction with this 'counting-oriented' contextualization, several students had trouble with making sense of the sampling procedure. Similar to the results of Brousseau et al. (2002), the students were not comfortable with the procedure of replacement. They considered it to be unfair in accordance with their contextualization, since one might count a piece that had already been counted. Being aware of this critical aspect, the introduction of Lesson 1B was adjusted to support a discussion that turned the students' attention to the proportions of the colors in the bag and to the meaning of a sample. As a result, the students were in a better position to benefit from the experiment and from the subsequent discussion where the investigator challenged them to discern aspects of chance by focusing on differences in percentages between the groups' samples.

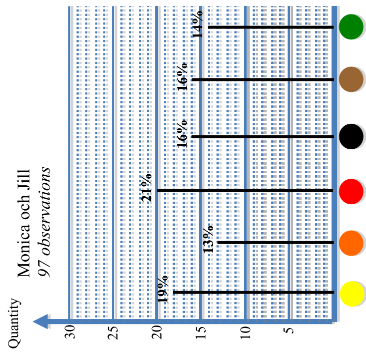
In order to compare and see the relationship between theoretical and empirical probability, the students need to develop an understanding of the part–whole relationship. From the previous line of reasoning, we can, on the one hand, claim that the students should enter probability experimentation with a full-fledged understanding of proportions and percentages in order to benefit from experimentation in probability. However, on the other hand, the teaching experiment described also shows that probability experimentation can serve to support the development of proportions and percentages. What we need to be clear about, however, is the questions we, as researchers or teachers, use to engine (Makar and Rubin 2009) reflection and discussion. In the present case, we have seen how we can get students to develop their thinking about probability by asking questions in the form of comparisons. A clear example of how such a question-form can have a positive impact on the students was when they changed from a part–part focus to a part–whole focus as they were challenged to compare the chances of getting a yellow from a bag of 200 pieces, containing 25 yellow, with a bag of 250 pieces, containing 26 yellow. Similarly, it became evident how the two comparison-oriented questions at the beginning of Lesson 2A and Lesson 2B challenged the students to discern central aspects of probability and promoted discussions on the relationship between theoretical and empirical probability.

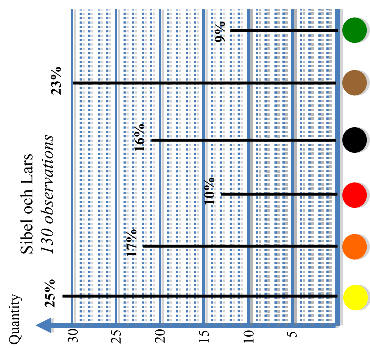
The differences between the groups' samples were not sufficiently distinct to challenge the students to consider sample size in the relationship between empirical and theoretical probability. Of course, on this issue, a computer environment would provide certain advantages (Pratt 2000; Stohl and Tarr 2002). In a computer environment, it is easy to repeat and to visualize the results of long-term random behavior in different ways. However, based on the present teaching experiment, one has to keep in mind that proportional thinking is not an easy thing for students to make sense of. On this issue, the present analysis indicates that there are challenges to the students when translating between the conceptual context of *part of a continuous quantity* (e.g., proportions visualized in a pie-chart) and the conceptual context of *part of a discrete quantity* (e.g., proportions visualized in bar-charts) (see, e.g., Charalambous and Pitta-Pantazi 2007). In connection with the reported difficulties students may have considering part–whole relationships and, perhaps especially in terms of part of a discrete quantity as was the case in the current experiment, there is reason to consider whether the students' tendency to increase the number of trials in a sample is based on a real understanding or if they are only using a technique which they adopt for the moment and which they would have difficulties translating to new situations. Our point here is not to choose between analog and digital learning environments. Rather, it is to encourage research on experimentation in the teaching and learning of probability to investigate the possibilities of developing learning environments which combine analog and digital experimentations in order to support conceptual understanding of probability. New digital devices such as iPads, mobile phones, and interactive whiteboards may help to support such combinations.

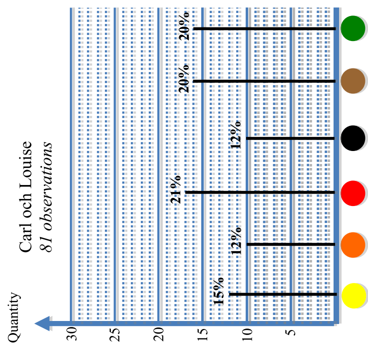
Appendix A


Quantity
Lisa and Eva
126 observations
30
25
20
15
10
5
9%
21%
18%
16%
24%
13%

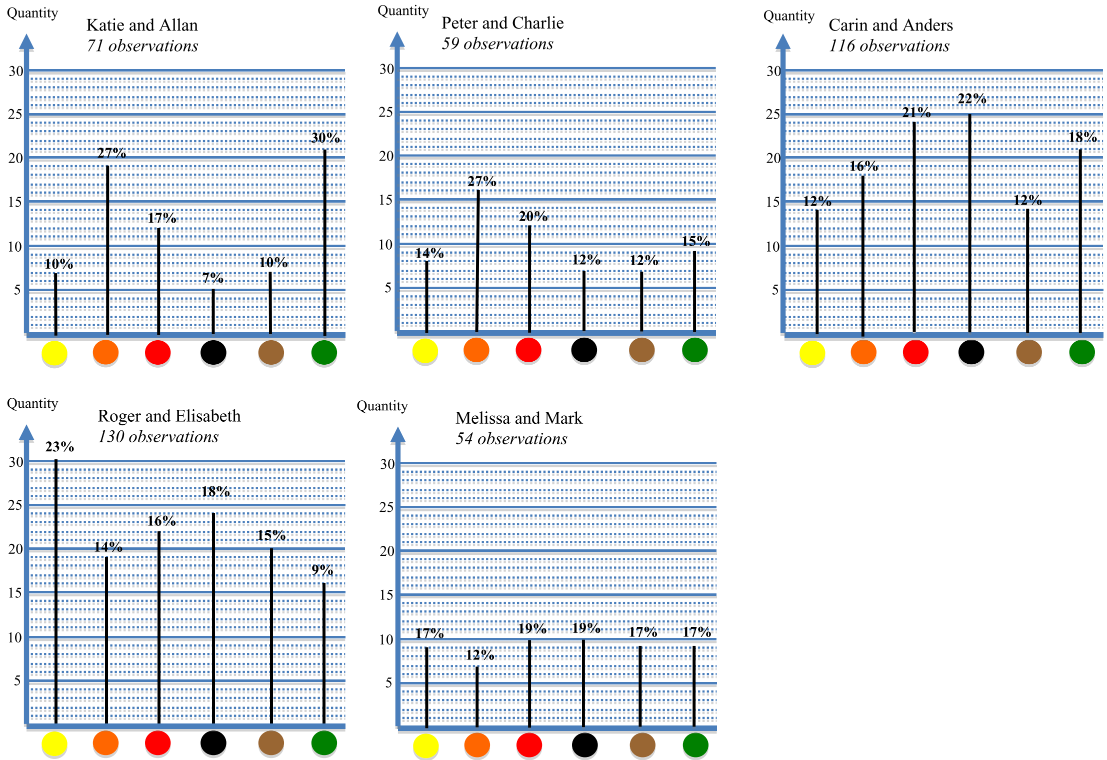

Quantity
Sara and Sergio
139 observations
30
25
20
15
10
5
14%
17%
14%
21%
17%
18%


Quantity
Monica och Jill
97 observations
30
25
20
15
10
5
19%
13%
21%
16%
16%
14%


Quantity
Sibel och Lars
130 observations
30
25
20
15
10
5
25%
17%
10%
16%
23%
9%


Quantity
Carl och Louise
81 observations
30
25
20
15
10
5
15%
12%
21%
12%
20%
20%

Appendix B


Quantity
Katie and Allan
71 observations
30
25
20
15
10
5
10%
27%
17%
7%
10%
30%
Quantity
Peter and Charlie
59 observations
30
25
20
15
10
5
14%
27%
20%
12%
12%
15%
Quantity
Carin and Anders
116 observations
30
25
20
15
10
5
12%
16%
21%
22%
12%
18%
Quantity
Roger and Elisabeth
130 observations
30
25
20
15
10
5
23%
14%
16%
18%
15%
9%
Quantity
Melissa and Mark
54 observations
30
25
20
15
10
5
17%
12%
19%
19%
17%
17%

References

Aspinwall, L., & Tarr, J. E. (2001). Middle school students' understanding of the role sample size plays in empirical probability. *The Journal of Mathematical Behavior*, *20*, 229–245.

Batanero, C., Henry, M., & Parzysz, B. (2005). The nature of chance and probability. In G. A. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 15–37). Media: Springer.

Borovcnik, M., Bentz, H.-J., & Kapadia, R. (1991). A probabilistic perspective. In R. Kapadia & M. Borovcnik (Eds.), *Chance encounters: probability in education* (pp. 27–71). Dordrecht: Kluwer Academic.

Brousseau, G., Brousseau, N., & Warfield, V. (2002). An experiment on the teaching of statistics and probability. *The Journal of Mathematical Behavior*, *20*, 363–411.

Charalambous, C., & Pitta-Pantazi, D. (2007). Drawing on a theoretical model to study students' understandings of fractions. *Educational Studies in Mathematics*, *64*, 293–316.

Chernoff, E., & Zazkis, R. (2011). From personal to conventional probabilities: from sample set to sample space. *Educational Studies in Mathematics*, *77*(1), 15–33.

Cobb, P. (1999). Individual and collective mathematical development: the case of statistical analysis. *Mathematical Thinking and Learning*, *1*, 5–43.

Halldén, O. (1999). Conceptual change and contextualization. In W. Schnotz, M. Carretero, & S. Vosniadou (Eds.), *New perspectives on conceptual change* (pp. 53–65). London: Elsevier.

Hawkins, A., & Kapadia, R. (1984). Children's conception of probability—a psychological and pedagogical review. *Educational Studies in Mathematics*, *15*, 349–377.

Horvath, J., & Lehrer, R. (1998). A model-based perspective on the development of children's understandings of chance and uncertainty. In S. P. Lajoie (Ed.), *Reflections on statistics* (pp. 121–148). Mahwah: Erlbaum.

Jones, G. A., Langrall, C. W., & Mooney, E. S. (2007). Research in probability: responding to classroom realities. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 909–956). Charlotte: Information Age Publishing.

Konold, K., Madden, S., Pollastek, A., Pfannkuch, M., Wild, C., Ziedins, I., Finzer, W., Horton, N., & Kazak, S. (2011). Conceptual challenges in coordinating theoretical and data-centered estimates of probability. *Mathematical Thinking and Learning*, *4*, 68–86.

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, *8*(1), 82–105.

Marton, F., Runesson, U., & Tsui, A. (2004). The space of learning. In F. Marton & A. Tsui (Eds.), *Classroom discourse and the space of learning* (pp. 3–40). Mahwah: Erlbaum.

Nilsson, P. (2009). Conceptual variation and coordination in probability reasoning. *The Journal of Mathematical Behavior*, *29*(4), 247–261.

Polaki, M. V. (2002). Using instruction to identify key features of Basotho elementary students' growth in probabilistic thinking. *Mathematical Thinking and Learning*, *4*, 285–314.

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

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

Pratt, D., Johnston-Wilder, P., Ainley, J., & Mason, J. (2008). Local and global thinking in statistical inference. *Statistics Education Research Journal*, *7*(2), 107–129.

Runesson, U. (2005). Beyond discourse and interaction. Variation: a critical aspect for teaching and learning mathematics. *The Cambridge Journal of Education*, *35*, 69–87.

Ryve, A. (2006). *Approaching mathematical discourse—two analytical frameworks and their relation to problem solving interactions*. PhD Thesis, Mälardalen University, Department of Mathematics and Physics.

Ryve, A., Nilsson, P., & Mason, J. (2011). Establishing mathematics for teaching within classroom interactions in teacher education. *Educational Studies in Mathematics*. doi:10.1007/s10649-011-9371-9. Advance online publication.

Saldanha, L., & Thompson, P. W. (2002). Conceptions on sample and their relationships to statistical inference. *Educational Studies in Mathematics*, *51*, 257–270.

Shaughnessy, J. M. (2003). Research on students' understanding of probability. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 216–226). Reston: The National Council of Teachers of Mathematics, Inc.

Stohl, H., & Tarr, J. E. (2002). Developing notions of inference with probability simulation tools. *Journal of Mathematical Behavior*, *21*(3), 319–337.

Säljö, R., Riesbeck, E., & Wyndhamn, J. (2003). Samtal, samarbete och samsyn: en studie av koordination av perspektiv i klassrumskommunikation. In O. Dysthe (Ed.), *Dialog, samspel och lärande*, Sweden, Studentlitteratur (pp. 219–240).

Wistedt, I., & Brattström, G. (2005). Understanding mathematical induction in a cooperative setting: merits and limitations of classroom communication among peers. In A. Chronaki & M. Christiansen (Eds.), *Challenging perspectives on mathematics classroom communication* (pp. 173–203). Greenwich: Information Age Publishing.

Investigating the Dynamics of Stochastic Learning Processes: A Didactical Research Perspective, Its Methodological and Theoretical Framework, Illustrated for the Case of the Short Term–Long Term Distinction

Susanne Prediger and Susanne Schnell

Abstract Our didactical research perspective focuses on stochastic teaching–learning processes in a systematically designed teaching–learning arrangement. Embedded in the methodological framework of Didactical Design Research, this perspective necessitates the iterative interplay between theoretically guided design of the teaching–learning arrangement and empirical studies for investigating the initiated learning processes in more and more depth. For investigating the micro-level of students' processes, we provide a theoretical framework and some exemplary results from a case study on students (in grade 6) approaching the distinction between short term and long term in the teaching–learning arrangement "Betting King".

In the last decades, a lot of research has been conducted on students' biases and misconceptions which seem to persist even after school education (overview in Shaughnessy 1992, pp. 479ff). Whereas most of these studies mainly focus on the *status* of (mis-)conceptions (as results of learning processes or as their initial starting points), we want to present a Didactical research perspective that complements these important studies by two dimensions: (i) the dynamic focus lies on specifically initiated learning *processes*. For this, (ii) a theoretically guided and empirically grounded *design* of teaching–learning arrangements (with restructured learning contents and concrete learning opportunities) and its underlying design principles are developed.

In Sect. 1 of this article, we present the didactical research perspective with its methodological framework of a process-oriented Didactical Design Research (Gravemeijer and Cobb 2006; Prediger and Link 2012). In Sect. 2, the perspective is exemplified by a report on six phases of research and (re-)design in a long term project on the distinction of short term and long term stochastic contexts for students in lower middle school. In Sect. 3, specific emphasis is put on the fourth phase, the investigation of the dynamics of stochastic learning processes on the micro-level. As the theory has been developed within these six phases, we present it together with

S. Prediger (✉) · S. Schnell
Institute for Development and Research in Mathematics Education, TU Dortmund University, Vogelpothsweg 87, 44227 Dortmund, Germany
e-mail: prediger@math.uni-dortmund.de

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_29,

each phase of the project (in Sects. 2 and 3). In Sect. 4, we conclude by extracting principles of a Didactical research perspective.

1 Methodological Framework

1.1 Didactical Design Research with a Focus on Learning Processes

The scientific work in mathematics education research is sometimes dichotomised by two different aims (that appear in Fig. 1 as both ends of the left vertical arrow): (i) Practical developmental work aims at *developing* general approaches and *designing* concrete teaching–learning arrangements for mathematics classrooms. (ii) Fundamental empirical research, on the other hand, aims at *understanding* and *explaining* students' thinking and teaching–learning processes.

To overcome this unfruitful dichotomy, more and more researchers advocate the general idea to join empirical research and the design of teaching–learning arrangements in order to advance both: practical designs and theory development. Under varying titles like "design science" (Wittmann 1995), "design research" (e.g. van den Akker et al. 2006; Gravemeijer and Cobb 2006), "design-based research" (e.g. Barab and Squire 2004) or "design experiments" (e.g. Brown 1992; Cobb et al. 2003; Schoenfeld 2006), programmes have been developed with this common idea (and variance in priorities).

In our research group, we follow the programme of Didactical Design Research as formulated by Gravemeijer and Cobb (2006), which seeks to combine the concrete design of learning arrangements with fundamental research on the initiated learning processes. Through iterative cycles of (re-)design, design experiment and analysis of learning processes, it focuses on both: (i) *design*: prototypes of teaching–learning arrangements and the underlying theoretical guidelines (that we call design principles) are created; (ii) *research*: an empirically grounded, subject-specific local teaching–learning theory is elaborated, specifying the following: the structure of the particular learning content, students' momentary state in the learning process, typical obstacles in their learning pathways, and conjectured conditions and effects of specific elements of the design (Prediger and Link 2012; Gravemeijer and Cobb 2006, p. 21).

As the research objects for Didactical Design Research are not only learning goals, contents and momentary states of learning, but also the teaching–learning *processes*, we need data collection methods that allow for these complex processes to take place and make them accessible for an (in-depth) analysis. For this, the method of design experiments is outlined briefly in the next section before introducing the research project on the distinction of short term and long term context as an example for the programme.

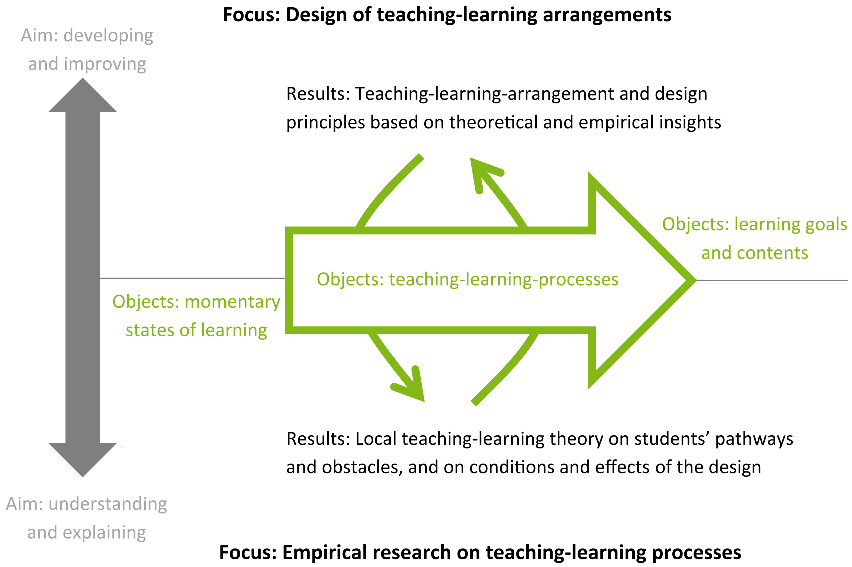


Fig. 1 Foci, aims, objects, and results of the Didactical Design Research

1.2 Design Experiments as Data Collection Method in Didactical Design Research

The intention of Didactical Design Research is to join design activities and empirical research by analysing processes that take place in a specifically constructed teaching–learning arrangement. The iterative process of designing and researching can start from conducting clinical interviews with students on their momentary state of learning in order to identify problems, prerequisites or conditions to specify the learning goals and/or give indications for the construction of the teaching–learning arrangement.

Data collection must therefore be optimized for investigating teaching–learning processes initiated within these arrangements. For this purpose, design experiments proved to be a fruitful method as they provide the following necessary characteristics (cf. Komorek and Duit 2004; Cobb et al. 2003): (a) Different from clinical interviews, design experiments initiate and support learning processes by adequate materials, activities and moderation of the teacher. Thus, the teacher can be a participating element of the design experiment which may influence the process through a methodologically controlled way of interaction (see Sect. 3 for an example). Accordingly, the situation is closer to normal in-classroom work on a specific topic than a clinical interview in which the interviewer is not supposed to influence students' thinking. (b) Furthermore, the method allows the observation of longer-term learning pathways by conducting several consecutive design experiment sessions. This can provide insights into sequences of applying, consolidating and deepening constructed knowledge. (c) Lastly, design experiments can be conducted in laboratory settings with small groups or pairs of students. This approach is fruitful to gain in-depth insights into individual, context-specific learning pathways, obstacles or individual prerequisites (cf. Komorek and Duit 2004). A laboratory setting is also suitable when the main aim is to specify and structure (possibly new) mathematical contents for improving curricula. However, before releasing a teaching–learning arrangement on a large scale, it is useful to investigate the ecological validity of the design and of the local teaching–learning theory in design experiments in classroom settings that are as natural as possible (e.g. with the regular teachers and normal resources) (Cobb et al. 2003; Burkhardt 2006).

The teaching–learning processes initiated in the design experiments are videotaped. The *data corpus* includes the videos, transcripts of selected video-sequences, all teaching materials, students' products and records of computer simulations.

The *data analysis* of the complex process data requires interpretative qualitative methods that are chosen according to the specific research interest in each phase of the process. We exemplify one analysis procedure in Sect. 3 for the in-depth analysis on students' learning process regarding the distinction between long term and short term contexts in the following section.

2 The Macro-Level: Six Phases of a Didactical Design Research Project on the Distinction of Short Term and Long Term Contexts

In order to illustrate the programme of Didactical Design Research, we present a project on the distinction between short term and long term stochastic context in grade 6 (age 11 to 13). Figure 2 gives an overview on the six phases of the project to be explained in the following sections.

The visual overview shows already that the close relation between design and research is a crucial principle for the programme of Didactical Design Research. Design and research are consequently related to each other and profit from each other: designing the teaching–learning arrangement was not only intended to provide students with better learning opportunities but also to create an environment for research that aims at generating a content-specific local teaching–learning theory where a gap in research was perceived. Vice versa, the results of the empirical research informed the (re-)design of teaching–learning arrangements so that teachers and students have direct benefits from the research.

2.1 First Phase: From an Empirical Problem to Specifying Learning Goals (2002–2005)

The project started from an observed problem of limited success of education in probability: Even adults who learned about probability in school rarely activate their

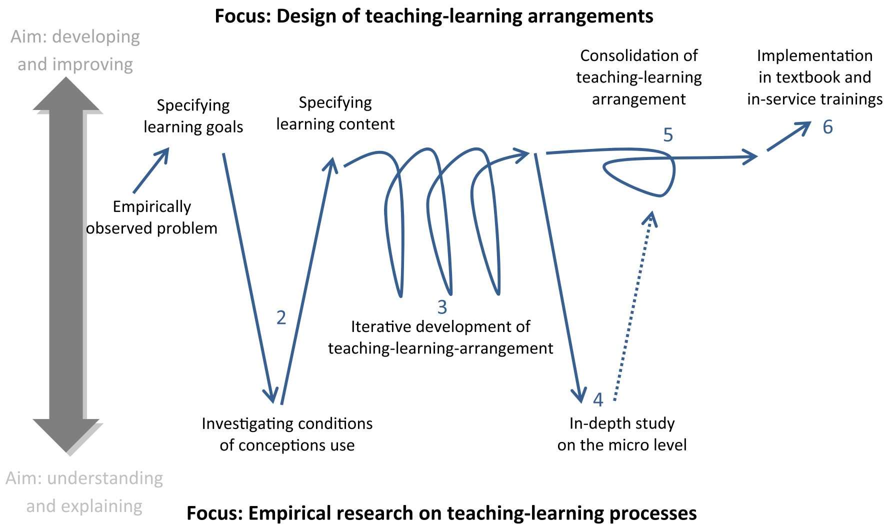


Fig. 2 Six phases in the long-term Didactical Design Research project on the distinction between short term and long term contexts

(existing!) stochastic knowledge when playing with random devices or betting on random trials in out-of-school contexts (Shaughnessy 1992, p. 465). They argue, for example, while betting on the sum of two dice, "Even if you can calculate that the eight is more probable, the dice don't show it, see! So I rather take my lucky number 12" (Prediger 2005, p. 33). In contrast, as long as students are in a probability classroom context, they do not hesitate to activate probability judgements for games or betting situations (Prediger 2005). But as long as probabilistic conceptions are restricted to classroom contexts, probability education fails to prepare for out-of-school contexts. From this problem, the overall goal for the instructional design was derived: *Enabling learners to activate elementary probabilistic conceptions context-adequately*.

A scientific design of teaching–learning arrangements requires a theoretical foundation, on which the specified learning goal can be justified and conceptualised. For this, we adapted the conceptual change approach on a constructivist background (Posner et al. 1982; Duit and Treagust 2003) and applied it to the specific learning content of probability and random phenomena (Prediger 2008).

Our conceptualisation of the term 'conception' follows Kattmann and Gropengießer (1996, p. 192) who refer it to all cognitive structures which "students use in order to interpret their experience" (ibid.). These cognitive structures are located on different epistemological levels of complexity, such as concepts, intuitive rules, thinking forms and local theories (Gropengießer 2001, p. 30 ff.). In line with the constructivist background, we understand everyday conceptions (e.g. the aforementioned 'lucky number') as important starting points for individual learning processes, even if they might not match the mathematical theory. The importance goes back to the constructivist position that active constructions of mental structures always build upon existing ones by accommodation to experiences with new phenomena, while the initial structures serve as "both a filter and a catalyst to the acquisition of new ideas" (Confrey 1990, p. 21). Thus, each design should take the initial conceptions seriously into account and develop them into the mathematically intended ones, as has been emphasised in conceptual change approaches (Duit and Treagust 2003).

This development does not always take place in terms of enabling students to *overcome* initial misconceptions and *substitute* them with mathematically appropriate ones (*vertical* conceptual change). In contrast, the goal of context-adequate activation of conceptions was specified within the theoretical framework of *horizontal* conceptual change (Prediger 2005, 2008), in which the instructional goal is to provide students with opportunities to *additionally* build mathematically sustainable conceptions and to enable them to consciously consider the context in order to *choose* which conception to activate (*horizontal* conceptual change). For reaching this goal in probability education, it is important that learners understand *when* mathematical conceptions about probabilities are applicable.

2.2 Second Phase: From Investigating Conditions of Students' Conceptions Use to Specifying the Learning Content "Distinction of Stochastic Contexts" (2004–2006)

The theoretical framework of horizontal conceptual change suggested an empirical investigation of the conditions under which students tend to activate adequate probabilistic conceptions (Prediger 2005, 2008). In clinical game interviews where children (about ten years old) were asked to bet on the sum of two dice, we could reproduce Konold's (1989) observation that many individual difficulties with decisions under uncertainty arise not only from problematic judgements on probability, but root deeper in "a different understanding of the goal in reasoning under uncertainty" (Konold 1989, p. 61). Konold could give empirical evidence for a predominant individual wish to predict single outcomes when dealing with instruments of chance, the so-called *outcome approach*. Similarly, the children in our interviews tried to explain the last outcome, drew conclusions from one outcome, tried to predict the next outcome, found evidence for the unpredictability of outcomes when outcomes did not follow the theoretical considerations etc. This outcome-approach could be reconstructed to be a main obstacle for the children to activate suitable probabilistic conceptions. But unlike Konold's results, our study could not reproduce the outcome approach as a *stable* phenomenon describing the behaviour of some children, but as *appearing situatively*. Students switched between the perspectives even without being aware of it: the same children could adopt a long-term perspective two minutes later, e.g. when considering a tally sheet with 200 outcomes. Hence, we

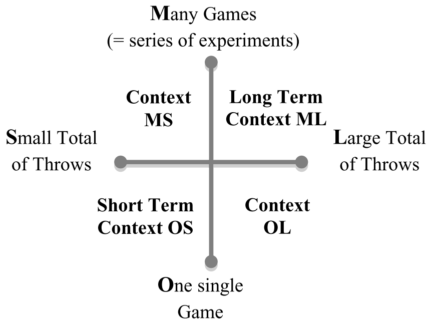


Fig. 3 Distinction of stochastic contexts

consider the so-called "misconception" of the "law of small numbers" (see Tversky and Kahnemann 1971) not to be wrong per se but only used within an inadequate domain of application. That is why the well-known distinction between short term and long term contexts turns out to be the *crucial background for context-adequate choices*.

Hence, the shift from a short term focused outcome-approach to a long term perspective on randomness (shortly called the stochastic context) is one crucial challenge for conceptual change. The horizontal view on conceptual change emphasises the need for individuals to become aware of the empirical law of large numbers as a condition for predicting outcomes in long term random situations by probabilistic terms.

In consequence, the construction of our teaching–learning arrangement was guided by the idea to allow children to recognise patterns in the long run (i.e. the stability of relative frequencies) and to make clear that the regularity described by probability does not apply to short sequences of random outcomes (Moore 1990, pp. 120–121; see also Konold 1989; Prediger 2008). For a mathematical–theoretical foundation of the design, the distinction of the stochastic context into the following two aspects proved to be fruitful (Fig. 3).

O–**M**. It is inherent in the construct of "regular pattern", even for deterministic phenomena, that it generalises from one experiment (for example, a game with dice) to a series of many experiments. Without considering many experiments, the existence or non-existence of regularities cannot even be discovered or stated. (Distinction **O**ne game—**M**any games, or in statistics **O**ne sample or **M**any (hypothetical) samples).

S–**L**. Moore (1990) characterises *random* phenomena as those phenomena where patterns only appear in experiments with many repetitions, for example, dice games with a large total of throws. For games with small total of throws, no (mathematically meaningful) pattern is visible. (Distinction **S**mall total of throws—**L**arge total of throws, or in statistics **S**mall sample—**L**arge sample.)

Fig. 4 The game 'Betting King'

2.3 Third Phase: Iterative Development of a Teaching–Learning Arrangement in Three Design Cycles (2005–2008)

Having specified the learning content as "distinction of short term and long term focus", we began the design of a teaching–learning arrangement by developing the game "Betting King" (Hußmann and Prediger 2009). This phase and all subsequent phases took place within the research context of the joint project KOSIMA that develops and investigates a complete middle school curriculum (Hußmann et al. 2011).

In accordance with Fischbein (1982), the main activity consists of guessing outcomes of a chance experiment by placing bets. In line with other designs for learning environments, this activity was embedded into a game situation (e.g. Aspinwall and Tarr 2001).

In the game "Betting King", the players bet on the winner of a race with four coloured animals (Fig. 4). A coloured 20-sided die with an asymmetric colour distribution (red ant: 7, green frog: 5, yellow snail: 5, blue hedgehog: 3) is used to move the animals.

For learning to distinguish stochastic contexts, students are guided to vary the total number of throws after which the winning animal is determined (**S**–**L**). It is materialised on the board by a throw counter (black token) and the STOP-sign which can be positioned between 1 and 40. The board game is followed up by a computer simulation that produces results of a game with the specified throw total up to 10 000 (Fig. 5). This allows students to find patterns in many games with the same total of throws (**O**–**M**) or to compare patterns between low and high totals of throws (**S**–**L**). Later in the learning situation, another mode is added to the simulation where a gradually growing throw total is provided to offer a dynamic perspective on the stabilisation of relative frequencies.

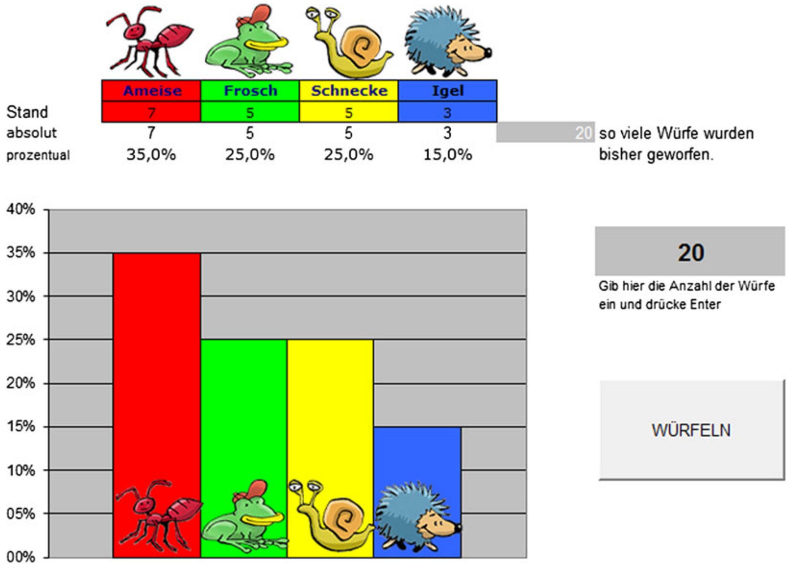


Fig. 5 Screenshot of computer simulation, here ideal distribution for a throw total of 20

The game is played in two variations: In the first game, students bet on the animal that will win after the specified total of throws (called "Betting on Winning"). Here, a merely comparative view on the chances of animals is sufficient, enabling the experience that a bet on the red ant (as the animal with the highest theoretical probability of $7/20$) is more likely to be successful for large totals of throws. In the second game ("Betting on Positions"), students are asked to predict the position on which each animal will land (between 0 and the pre-set throw total), i.e. to bet on the frequencies. In order to focus on *relative* deviances instead of absolute ones, the bets of the two players are compared, and the bet closer to the actual position per animal gets a point. The need to find good betting strategies motivates the exploration of patterns of relative frequencies in the long run.

The teaching–learning arrangement consists not only of the game Betting King, but also of a sequence of initiated activities and questions and of instructional means like the computer game, visualisations, and records that focus students' attention. While playing the game variations themselves allows experiences with the phenomena of stabilising relative frequencies, the phase of systematic investigation of patterns focuses more consequently the attention. After these investigations, a phase of institutionalisation and systematisation is initiated, then a transfer to other random devices and some training for consolidation and deepening the gained ideas (Prediger and Hußmann 2014).

This complex arrangement is the product of three cycles of design research. We now briefly describe the process of the iterative interplay of design experiments, analysis and redesign:

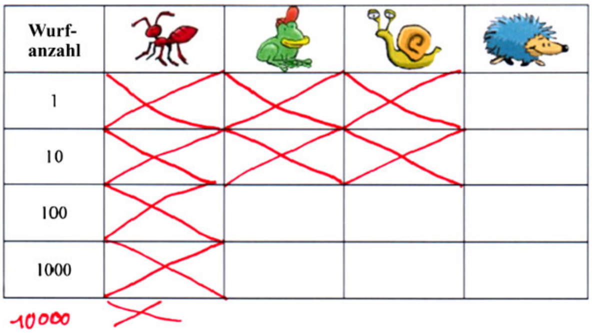


Fig. 6 Filled in summary table of the students Hannah and Nelly

First Cycle The first ad hoc trial in the first authors' regular mathematics class ($n = 22$, grade 6, age 11 to 13) helped to develop the ideas and rules of the game and to define requirements for the computer game so that the intended learning goals could be reached. For example, the STOP-sign was introduced for materialising the difference between small and larger total of throws on the board.

Second Cycle The second design experiment in laboratory settings gave first insights into the learning pathways and in productive sequences of activities and questions ($n = 2 \times 2$, grade 4, Plaga 2008). After these design experiments, the game was complemented by reflection tasks and activity sheets. The phase of *playing* the game was explicitly separated from *investigating* the patterns: "In this task, you can now investigate the game in more detail. This will help you to find a good strategy for betting." While playing, students often focus on one single game with a small or large total of throws and sometimes avoid reflecting betting strategies. In contrast, the investigation of patterns focuses on many games, which facilitates the acquisition of the long term perspective. The investigations were structured by records with pre-defined throw totals such as 4×1, 4×10, 4×100, $4 \times 1\,000$ and the suggestion to compare the results for each throw total with each other.

Third Cycle The third design experiment in four classes ($n \approx 100$, grade 5/6) with their regular teachers helped to ensure the feasibility of the arrangement under classroom conditions and to evaluate the learning success. The empirical analysis of classroom learning processes in Prediger and Rolka (2009) showed that most students could find adequate betting strategies and learned to differentiate between long term and short term contexts. However, weaker students experienced difficulties coping with variation for small total of throws. For example, Amelie wrote when she compared the frequencies of wins in a table:

> On the table, it catches my eye that the ant has more points than the others. But the other players (snail, frog, and hedgehog) also won sometimes. The ant is not the fastest because the others also won sometimes. (Prediger and Rolka 2009, p. 64)

We interpreted Amelie's written utterance in the way that she had not yet developed a language to distinguish between purely ordinal judgements (the red ant is more likely to win than the other animals for each total of throws) and judgements regarding the empirical law of large numbers (the red ant wins more securely for 1 000 throws than for 10). Consequently, the distinction between 'good bet' and 'more secure bet' was introduced into the arrangement. After filling in the summary table showing if an animal has won in any game with a certain throw total (cf. Fig. 6), students were asked: "For making good bets, you look for an animal with the highest chance to win. But you can never be completely sure. For some throw totals, though, you can be surer than for others. Find good throw totals by comparing results from all games: For which throw total can you bet more securely?"

This is one among many examples where the investigation of students' thinking led to important changes in the learning arrangements since it could clarify the key points of the learning content (Prediger 2008). However, the classroom observations only allowed a rather rough look at the learning processes. Deeper insights into students' pathways towards the distinction between good and more secure bets as facet of the intended differentiation of stochastic contexts needed a deeper analysis as conducted in the fourth phase.

2.4 *Fourth Phase: In-Depth Process Study on the Micro-Level: Investigating the Dynamics of Stochastic Learning (2009–2012)*

On the base of a functioning teaching–learning arrangement that enables students to reach the learning goals (see third phase), a further series of design experiments was conducted in a laboratory setting (Komorek and Duit 2004) by the second author within her PhD-project (Schnell 2013, supervised by the first author). The new empirical study on the micro-level aimed at deepening the local theory on students' individual construction of distinctions between short term and long term contexts by addressing patterns and deviations in the empirical approach. The design experiments were conducted with nine pairs of students (grade 6, age between 11 and 13), with four to six consecutive sessions each.

The micro-level analysis of individual learning pathways required further elements in the theoretical framework and a method of data analysis as will be further elaborated in Sect. 3. It also offers some insights not only into typical conceptions but also into their networks and into the processes how they are gradually constructed and used with regard to the situational and stochastic context.

Although the fourth phase of design experiments also informed some changes in the learning arrangement, the main emphasis was on the elaboration of the empirically grounded local theory.

2.5 Fifth Phase: Consolidation of the Teaching–Learning Arrangement (2010–2011)

For evaluating the usability of the developed teaching–learning arrangement under field conditions, a final, fifth design cycle was conducted in two classes ($n \approx 54$, grade 6). In this cycle, material and hints for the teachers' handbook were collected.

2.6 Sixth Phase: Implementation in a Textbook and in in-Service Trainings (from 2012)

The resulting teaching and learning materials are included into the innovative textbook series "Mathewerkstatt 6" (Prediger and Hußmann 2014). Its dissemination to schools is supported by a widespread series of in-service teacher trainings within the KOSIMA-teacher-network (Hußmann et al. 2011).

3 The Micro-Level: In-Depth Process Study

Building upon the findings of the analysis of the first three iterative design cycles, an in-depth process study was conducted in order to gain more fine-grained insights into students' learning pathways. The main goal was to generate a subject-specific local teaching–learning theory on the development of conceptions about the empirical law of large numbers and the distinction of short and long term contexts. The part of the study presented here was guided by the following research questions:

1. How do students successively construct the distinction of short and long term contexts within the teaching–learning arrangement 'The Betting King'?
2. Which conceptions are developed and how are these related to each other?

The study builds upon the conceptual change approach to conceptualise the development of conceptions on a macro-level. However, the framework had to be complemented by more fine-grained theoretical conceptualisations (Sect. 3.1) and by adapted analysis procedures (Sect. 3.2). Section 3.3 presents a new case of the large empirical data (analysed more broadly in Schnell 2013).

3.1 More Fine-Grained Theoretical Conceptualisation of Conceptions: Constructs and Its Elements

Within the preceding cycles, the conceptual change approach proved to be useful to describe the general structure of learning processes in regard of horizontal and vertical developments. For this first approach, intended and initial conceptions were conceptualised in an epistemologically wide sense without further distinctions (see Sect. 2.1).

For complementing this theoretical framework on the micro-level (Schnell and Prediger 2012), we adapted the notion of '*constructs*' as the smallest empirically-identifiable units of conceptions from the theoretical model 'abstraction in context' (Schwarz et al. 2009). These constructs are expressed in propositions, for example, ⟨Red ant is a good bet as it just won twice in a row⟩. Conceptions are then seen as webbings of constructs, i.e. they can consist of several constructs and their specific relations between each other (similarly in diSessa 1993). A theoretical adaptation of the notion of Schwarz et al. (2009) was necessary for grasping the horizontal dimension of conceptual change in which idiosyncratic conceptions are considered to be legitimate building blocks: Whereas Schwarz et al. (2009) mainly consider mathematically intended or mathematically (partially) correct constructs (Ron et al. 2010), we consider all individual constructs, being in line with mathematical conceptions or not.

For the empirical reconstruction of individual constructs, the theoretical model of abstraction in context provides an instructive methodology by means of three observable epistemic actions in a so-called *RBC-model*: recognizing (R), building-with (B) and constructing (C). An epistemic action of *constructing* is defined as (re-)creating a new knowledge construct by *building with* existing ones. Previous constructs can be *recognized* as relevant for a specific context and can be used for *building-with* actions in order to achieve a localised goal. The first step of our data analysis used the RBC-model for the reconstruction of constructs by identifying epistemic actions and their subjects. After this basic interpretation, our data analysis went further to describe the combinations of existing and the emergence of new constructs in more detail in order to gain insights into the microprocesses of conceptual change.

For the purpose of reconstructing these processes in depth, we followed diSessa (1993) and Pratt and Noss (2002) in their assertion that constructs cannot only be described by their *proposition*. In addition to the propositions, we took into account the construct's *function* for the individual, e.g. description of a pattern, explanation of the pattern, prediction of the results.

Furthermore, it proved necessary to conceptualise the individual scope of applicability. Whereas Pratt and Noss's (2002) notion of 'contextual neighbourhood' is defined as a repertoire of circumstances in which a construct can be used, we define the 'context of a construct' as the context in which the construct is created and/or used and distinguish between stochastic and situational contexts: the relevant *stochastic contexts* have been introduced in Sect. 2.2 in Fig. 3: Short Term Context OS (one game with a small total of throws), Context MS (many games with a small total of throws), Context OL (One game with a large total of throws), and Long Term Context ML (many games with large total of throws).

Whereas this distinction of stochastic contexts has to be constructed during the learning process, learners first distinguish between other contextual circumstances that we call *situational contexts*, for example, representations (numerical data table, bar charts, percentages), sources (empirical data, theoretical ideas, alternative ideas) or external settings (chance device, computer simulations vs. board game etc.). All elements of a construct are subject to negotiations during the learning process and can be revised, restricted or broadened in order to accommodate new experiences.

Table 1 Elements of a construct—an example

Element	General description	Description for the example ANT-GOOD-construct
Proposition (*prop*):	What is the proposition stated in the construct?	⟨Red ant is good as it just won twice in a row⟩
Stochastic context (*con*):	To which stochastic context does the construct refer? (ML, MS, OL, OS)	MS—Two games with a throw total of 20 and 24
Relevant parts of situational context (*sit*):	Which parts of the situational contexts are relevant for the constructing?	empirical database
Function (*fct*):	What is the construct used for? (e.g. description of a pattern, explanation of the pattern, prediction of the results)	description of a pattern, prediction

In Table 1, these elements are exemplified by the following scene from one of the design experiments with the students Hannah and Nelly (age 12 and 13) that will be further analysed in Sect. 3.3. In this scene (after 15 minutes of the first experiment session), the girls had been playing the game 'Betting on Winning' twice without using a record yet.

Played games so far
*(HN) indicates the bet of the children, (Iv) the interviewer's bet, ** marks the winning animal*

Transcript line	*Game number*	Throw total	Red ant	Green frog	Yellow snail	Blue hedgehog
306	*#1*	24	*9* (HN)	4	6 (Iv)	5
380	*#2*	20	*7*	3 (HN)	6	4 (Iv)

Though the girls cooperate in deciding about the throw totals and their bets, they had not yet given any reason for their betting decisions. The third game is about to take place and the throw total was set to 14.

404	H	(*places STOP-sign on field 14*)
405	N	Okay, which one do we bet on?
406	H	You decide.
407	N	Ant, it just won.
408	H	It won twice in a row already.
409	N	Yeees, it is good!
410	Iv	I'll take the snail. Okay, who begins?

ANT-GOOD
⟨Ant is good as it just won twice in a row⟩

Sit: empirical Con: MS Fct: pattern

In this scene, the girls develop a construct that we called ANT-GOOD (see right column next to the transcript). It summarises the children's experience in the previous games; its construction is thus situated in an empirical situational context. In line 407, Nelly points to the last game while Hannah focuses on both previous games, emphasising the addition "in a row". Thus she makes an explicit statement taking more than one game into account, so we dared—to this early moment—to code the stochastic context MS, i.e. as **M**any **S**mall games. Lastly, the function of the construct in the situation is to describe a pattern which is used to make a prediction for the next game.

3.2 Methods for Data Collection and Data Analysis

For data collection, the design experiments were conducted in a laboratory setting (Komorek and Duit 2004) with four to six sessions of 45 to 90 minutes each. The sample consisted of nine pairs of students of grade 6 (age 11 to 13) in a German comprehensive school ($n = 18$). The characteristics of a laboratory setting required some modifications of the teaching–learning arrangement: Since the laboratory setting lacks the dynamics of typical classroom interaction (exchange of findings and ideas between groups of students), the students needed more time for exploring data and gathering their own empirical base for generating hypotheses. As the research interest was on the development of conceptions, a deeper insight into students' reasoning was needed so that they were given more opportunities to verbalise or write down their ideas. The interviewer's interventions were guided by an intervention manual defining the sequence of tasks and the probing questions. The interventions were designed for minimal help, however, the continuation of the process was guaranteed by predefined interventions for anticipated obstacles. For example, UNEQUAL COLOUR-DISTRIBUTION is a crucial construct in order to give meaning to the observed patterns such as ⟨the red ant is the animal winning most often⟩. If the students do not discover the unfair distribution until having filled in the summary (cf. Fig. 6), the interviewer gives the advice to look closely at the die.

Nearly 40 sessions were videotaped in total, and the complete nine sessions of two pairs of students as well as selected sequences of other pairs were transcribed. The data corpus also included records of computer simulations and written products.

In the *first step of qualitative data analysis* of the transcripts, an adapted form of RBC-analysis procedure was conducted. The outcome of the analysis was a long list of individual constructs represented by propositions for all interviews (e.g. 101 constructs for one pair of students). The RBC-analysis also generated a first draft of the constructs' connections, given by the epistemic actions *constructing* and *building-with*.

In the *second step of data analysis*, these constructs were analysed according to their functions and contexts. The *third step of data analysis* dealt with identifying what happens in building-with processes, i.e. when previous constructs are developed or put in relation. We reconstructed all networks of constructs emerging in the design experiments of the completely transcribed pairs. By an interpretative approach, we reconstructed the nature of the relations between different constructs with respect to the modification of elements.

Due to page limitations, the following section can only present a part of the generated local theory. Other extracts are published in Schnell and Prediger (2012) and Schnell (2013).

Transcript line	*Game number*	Throw total	I bet on ...	Fastest animal	Points
500	*#4*	1	Ant	snail	0
540	*#5*	2	Hedgehog	snail, frog	0
602	*#6*	5	Snail	hedgehog, ant	0
651	*#7*	10	Frog	frog, snail	1
783	*#8*	20	Ant	frog	0
				Total points:	1

Fig. 7 Excerpt of Hannah and Nelly's record (first set of logged games, two left columns added by authors)

3.3 Case Study of Hannah's and Nelly's Distinction Between 'Good Bet' and 'Secure Bet'

In order to show what kind of insights on the micro-level can be reconstructed by this theoretical and methodological framework, we show spotlights of the case study of Hannah and Nelly's learning pathway when they successively construct the distinction between "a good bet" and "a secure bet". Thus the focus lies on the development of the stochastic context, whereas other constructs and the development of the construct-elements "situational context" and "function" are out of the focus here.

The 12 and 13 years old girls Hannah and Nelly approach the design experiment and laboratory situation openly; they work cooperatively and discuss their diverging opinions vividly. The first extract shows how they construct the context-distinction O–M, i.e. between one and many games.

In their first construct ⟨Ant is a good bet as it just won twice in a row⟩ ANT-GOOD (see transcript line 408 after game #2 above), the girls refer to the two games played so far (plus the on-going game in which red ant is leading), thus they take explicitly more than one single game into account ("in a row").

After this (in lines 436–1281, not printed), they start to play with the record in Fig. 7. It shows that they do not consequently use the ANT-GOOD construct for betting in game #4–#8 (in lines 436–795). Before starting the second record for games #9–#14, the following dialogue takes place:

796	N	We bet on frog.
797	H	Man! But frog won three times. Snail won three times, too. Ant once.
798	N	(*reads on record*) Fastest animal. We don't know that yet. I bet we get one point again (*points to "total points" at the bottom of the set*)
799	H	Hey, let's write down all of the [bets] already.
800	N	No.
801	Iv	What would you write down?
802	H	Frog, frog, frog, frog, frog. Okay, snail, frog.

TOTAL-WINS
⟨The number of total wins indicates a good animal to bet on⟩

Sit: empirical	Con: MS	Fct: pattern/ prediction

In line 797, Hannah offers a construct in which she compares the overall numbers of wins in all recorded games. The use of the conjunction "but" might be a reaction to the previous game #8 with a throw total of 20 in which both girls were convinced of their bet on the ant, but frog was the winner. Thus, she could hereby indicate that she has found a new strategy. In game #8, frog was clearly leading from the beginning. This could possibly influence the prediction Hannah makes in line 801, which doesn't seem to match to the fact that the absolute numbers of wins is the same for frog and snail. It is remarkable that in line 799, she suggests predicting all five upcoming games at once, which could indicate that she is already convinced of her new construct TOTAL-WINS.

Nelly refuses this suggestion in line 800. Taking a long shot, this could be due to her establishing focus on the previous game which might have led to her bet on frog in 796. But this construct is only made explicit about 15 minutes later after game #18:

1294	N	Yes, well… we almost always take the one that won before.
1295	H	And that always wins.
1296	N	And that wins more often.

LAST-WINNER ⟨The Animal who won the last game is a good bet⟩		
Sit: alternative/ empirical	Con: OS	Fct: pattern/ prediction

The construct LAST-WINNER is then tested in the course of the following games and finally written down as a common betting strategy ("We always take the animal who won in the last game", lines 1297–1335, not printed). Here, the focused stochastic context is clearly the last single game which is used as a predictor for the next game. The construct LAST-WINNER might have emerged since TOTAL-WINS could not give clear results: in the first 20 games, red ant and yellow snail won equally often.

However, the girls seem conscious of the limits of the LAST-WINNER's power of prediction. When the interviewer questions the LAST-WINNER strategy by pointing out that it sometimes doesn't work, Hannah answers:

1524	H	We said you win <u>more often</u>. We didn't say you always win.

This is possibly a reference to Nelly's statement in line 1296. Hannah's reasoning might be: the bets are results of chance, so that they are not completely predictable, but LAST-Winner with its focus on single games seems to be more successful than TOTAL-WINS with the focus on many games.

When later two animals win, LAST-WINNER is questioned and is built with the TOTAL-WINS construct into a combined construct:

1755	Iv	What will you take in the next round, now that two animals won?
1756	H	(*whispers*) Ant (*says loudly*) The ant.
1757	N	(*whispers*) Snail.
1758	H	Ant won more often than the snail.
1759	N	That's right. But the snail caught up now.
1760	H	But ant has one, two, three, four, five, six. And snail has only one, two, three (*points to games on record*).
1761	N	Ok, ant. (. . .)

LAST&TOTAL
⟨When two animals won in the last game, the animal with more wins in total is the better bet for the next game⟩

Sit: alternative/ empirical	Con: O + MS	Fct: pattern/ prediction

By this LAST&TOTAL construct, the perspective on the last one game and on a series of many games are combined. When the prediction by looking at the last game is unclear, the total number of throws is taken into account for predicting the next game. This combined construct emerges right before the students start to play with a throw total of 100, thus it is constructed for games with a small total of throws.

To sum up so far, Hannah and Nelly's learning trajectory diverges from the intended pathway, as they haven't identified the red ant as a good bet (neither empirically nor theoretically based) but found a deviant individual good bet based on the last outcome combined with a comparison of total wins. Although Hannah's statement in line 1524 about the quality of LAST-WIN could be interpreted as awareness that this seemingly good betting strategy is not yet a secure one, the distinction between single games and many games doesn't seem to have been constructed explicitly, yet. The notion of a "good but not secure" bet is developed already when looking at one or many games without taking the throw total in regard, yet.

The situation is solved when the girls discover the colour distribution and are then enabled to relate it back to their empirically based LAST&TOTAL strategy:

2192	N	[explains chances according to colour distribution] Blue, you don't have it that much. That means, when you get blue by chance. . . and then you write down blue as it was our strategy, then you probably won't get blue again.

ONE-GAME-ALL-POSSIBLE
⟨In one game, blue can win by chance but you probably won't get it again⟩

Sit: empirical	Con: O	Fct: pattern/ evaluation

Nelly evaluates her previous constructs (LAST&TOTAL or just LAST-WIN) and points out a problem: in a single game, an animal with a low chance such as the blue hedgehog could win. In a series of games though, it is least likely that it wins. Therefore, the single game is (sometimes) a bad predictor of the next winner. Thus, she now makes the distinction between one and many games explicit, but without mentioning whether it refers to small or large throw totals.

In the following extract, both girls construct the context-distinction S–L, i.e. between small and large totals of throws.

Even though creating a construct for finding an individually as good perceived bet (LAST&TOTAL), the context-distinction between small and large throw totals has not developed, yet. This takes place when the girls work on the record that has room for 16 games with pre-defined throw totals 1, 10, 100, 1 000, each four times (lines 1634–1870, not printed). In this sequence, the girls develop a confidence in the ant as the fastest animal and make first experiences of differences between small and large throw totals.

Then the girls are guided to compare all games' results with hindsight to which animal has won at least one time for which throw total (cf. Fig. 6 for Hannah's and Nelly's filled in table; lines 1871–2306, not printed). At this point, the unequal colour distribution is finally discovered (COLOUR-DISTRIBUTION, lines 2146–2186, not printed) with the help of the interviewer and both students explain the chances of each animal correctly. After this, they start working on the summarising question:

2205 Iv (*reads*) For which throw total can you bet more securely? And why?

2206 N I would say. So honestly I would write down it doesn't—

> **NO-THROW-TOTAL**
> ⟨The throw total doesn't matter for betting⟩
> Sit: empirical Con: all Fct: pattern

2207 H Yes?

2208 N Well, the number doesn't matter really. Well of course, for low numbers such as one or two, the chance isn't as big, because…

2209 Iv A chance for what do you mean?

2210 N Yes, well, to win. Because when… there are all—er, what do you say? (*laughs*) Colours.

2211 Iv Colours?

2212 N On the die. And for one, you can't really say who won, yet. That is why—So when you roll it more often, such as 20 times or such. Then you get a… more definite result, I think.

> **ALL-WIN**$_{1,2}$
> ⟨For a total of throws of 1 or 2, all colours have a chance to win⟩
> Sit: empirical Con: MS Fct: pattern

2213 Iv What do you mean by definite? Well, what do you mean?

2214 N Yes, I mean when… more often I say. For the first—for one, I say, blue won.

2215 Iv Mhm.

2216 N By chance! For two… no idea—green won. And for five, red can win already. Or something, right? And when you have 20, there are more chances to throw it more often.

> **BEST-ANT**$_{5,20}$
> ⟨For throw totals such as 5 and 20, chances for the ant are higher⟩
> Sit: empirical Con: MS Fct: pattern

This is the first time in the whole session that the students' attention is explicitly directed to the throw total. Thus Nelly's first reaction in line 2206 seems rather spontaneous. Her first construct is that the throw total doesn't matter at all, for which she promptly creates an exception concerning the stochastic context of throw totals of 1 or 2. Thus, NO-THROW-TOTAL seems to get restricted only to games with a throw total of more than two. By constructing the complementary constructs ALL-WIN$_{1,2}$ and BEST-ANT$_{5,20}$ in lines 2212 to 2216, Nelly invents a distinction between the throw totals in hindsight of the security of the ant's chance for winning. The throw total 5 seems to serve as a borderline case between ALL-WIN and ANT-BEST. Here, Nelly is explicitly constructing that there *is* a difference between small and large numbers of throws. She has not yet built a correct idea of *where* a border between small and large numbers can be adequately drawn.

The border is shifted in the next sequence when addressing the visualisation in the summary table (Fig. 6). Looking at the table, Nelly points out that red ant is the winning animal for a throw total of 100 and 1 000 "because it has more chances on the die because it is more often on it" (line 2222; BEST-ANT$_{100,1\,000}$). Nelly uses the construct COLOUR-DISTRIBUTION in a comparative way, describing the advantage of the red ant as an explanation for BEST-ANT$_{100,1\,000}$. The girls then start thinking about how to write down an answer to the questions: "For which throw total can you bet more securely? And why?".

2284	N	We believe—No I would write it like that: We believe for the ant, chances to win aren't so big for 1 to 10. Uh, not as big as for 100 to 10 000.
2285	H	Yes.
2286	N	To win.
2287	Iv	Mhm.
2288	N	Yes?
2289	H	As we think here (*points to summary, probably throw total 1 and 10*) are other animals next to it [the red ant] and here (*probably points to throw totals 100, 1 000 and 10 000*), there is no one else [other than the red ant].
2290	Iv	Right (...)

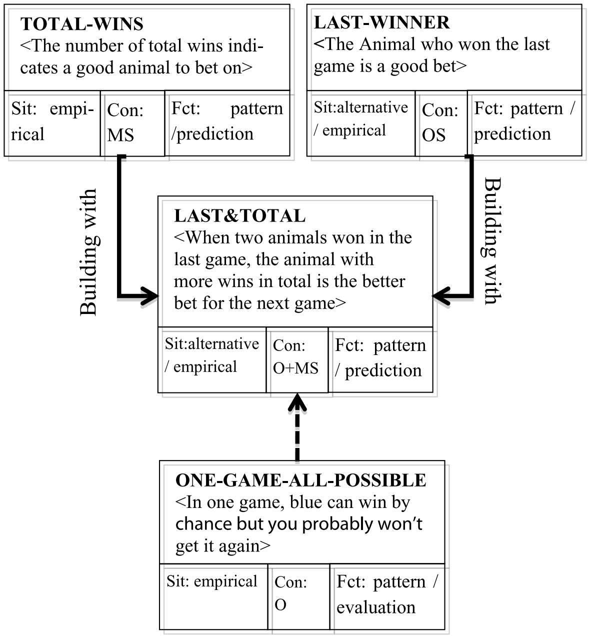
ANT-SMALL&LARGE
⟨Chances for red ant to win aren't as big for 1–10 as they are for 100 to 10 000, because there, it is the only one to win⟩

Sit: empirical Con: MS + L Fct: pattern

Nelly's statement in line 2284 gives insight how she is building with notions of good versus secure bet. Saying "chances to win aren't so big for 1 to 10" (line 2284) is put into relation to the high throw totals by adding "not as big as for 100 to 10 000". This statement might include that chances for the red ant are still high for small throw totals (which was maybe part of the construct ANT-BEST$_{2,5}$ as mentioned in line 2212) and this makes the comparison with the long term perspective necessary. She then correctly backs up her initial statement by relating it to the summarised empirical findings on the table in line 2289. Concluding, this statement could be understood as a relation between the constructs ANT-BEST$_{100,1\,000}$ and ALL-WIN$_{1,2}$, if the scope of applicability of the latter was broadened from only 1 and 2 to 1 to 10.

To sum up, the excerpts from Hannah and Nelly's learning pathway show how complex the processes of developing constructs are. The girls first develop the dis-

Fig. 8 Finding a **good bet** (distinction of **O**ne vs. **M**any games)



tinction of one versus many games (Fig. 8). Because finding a good bet can be hard for games with small total of throws (as the relative frequencies vary a lot), they use a combined strategy (LAST&TOTAL) of looking at a series of games (ANT-GOOD and TOTAL-WINS) and single previous results (LAST-WIN).

Only when the colour-distribution is discovered to be unfair, they evaluate their previous constructs and realise how badly single games serve for deriving a good bet of the next winning animals, because every animal can win "by chance" (ONE-GAME-ALL-POSSIBLE).

While the distinction of one versus many games establishes the notion of a good bet (in a comparative perspective), the distinction of small versus large total of throws gives insight into when a bet is secure (cf. Fig. 9). When the girls' focus is guided to the throw total, Nelly starts by a statement of not taking the throw total into account (NO-THROW-TOTAL) which is in line with their previous behaviour. She then promptly revises her idea by making exceptions for very low throw totals such as 1 and 2 and declaring that every animal can win here ($\text{ALL-WIN}_{1,2}$), while for throw totals such as 5 and 20, the ant has a higher chance to win ($\text{BEST-ANT}_{5,20}$). By mentioning these two constructs directly after each other, she builds complementing constructs for small and individually as large perceived throw totals (micro-developments like building complements are discussed in Schnell and Prediger (2012) in more detail). The material provided by the learning arrangement then supports her in making a distinction between small throw totals such as 1 and 10 and large throw totals such as 100 and 1 000 ($\text{BEST-ANT}_{100,1\,000}$) and relating them correctly to the ant's less or more secure chance of winning (ANT-SMALL&LARGE).

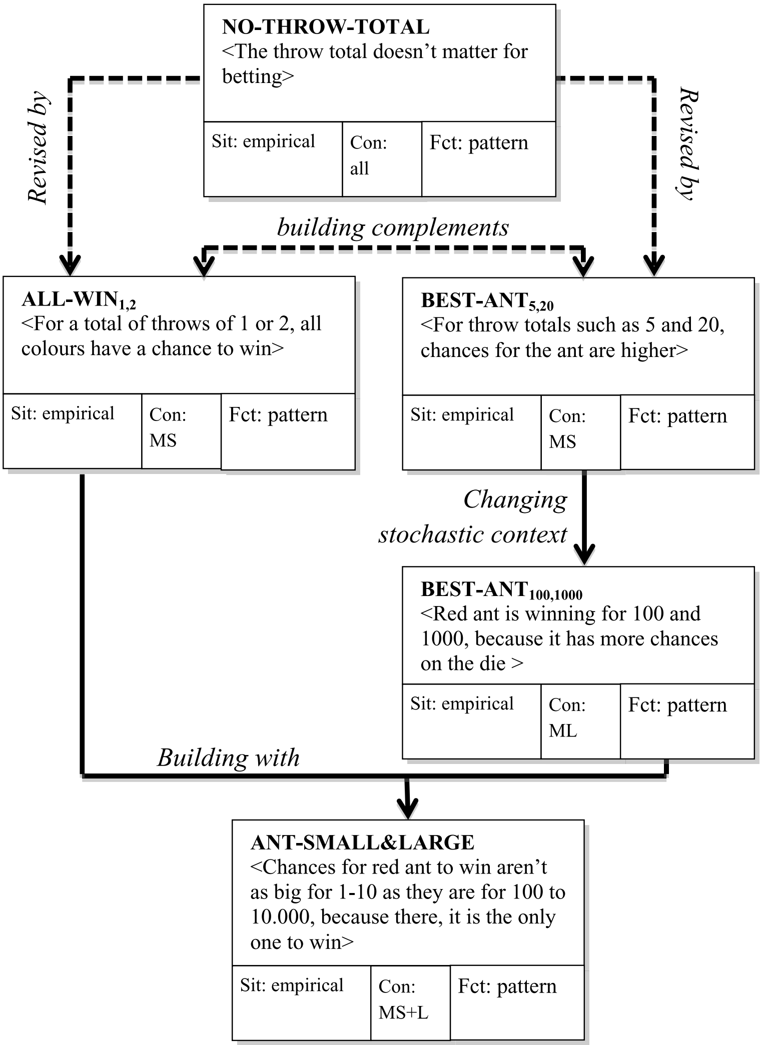


Fig. 9 Finding a **secure bet** (distinction of **S**mall vs. **L**arge total numbers of throws)

3.4 Conclusions

The analysis of the learning pathway on the micro-level allows fine-grained insights into the complex processes of developing a distinction between short and long term contexts. Even though this distinction was a crucial aim of the learning arrangement and supported by focussing record, tables and tasks, the results on the level of constructs show how complicated and interwoven the development is. This also highlights the advantages of researching learning pathways in laboratory settings: Fragile constructs and strategies can be traced within a setting with a reduced complexity. It allows reconstructing individual learning pathways by also giving room and time for verbalising and eventually building with volatile constructs. In classroom settings, these might be overlooked due to many influences and dynamics between students (and teachers).

Summarising, the presented excerpts from the case study illustrate the local theory of how the distinction of short term and long term contexts can develop. In the larger study (Schnell 2013), these findings are compared with other pairs of students, establishing a powerful framework for describing and understanding learning pathways within the learning arrangement "Betting King" on a micro-level.

4 Looking Back: Principles of a Didactical Research Perspective

The presented Design Research project on the distinction between short and long term contexts is a long-term project within the programme of Didactical Design Research. Although sketched only very roughly, we intended to exemplify four characteristics that are crucial for this programme (van den Akker et al. 2006; Gravemeijer and Cobb 2006; Prediger and Link 2012):

4.1 Use-Inspired Basic Research in Pasteur's Quadrant

As Burkhardt (2006) and others have emphasised, Design Research is equally oriented at basic research with fundamental understanding and at considerations of use for mathematics classrooms, it is hence located in the so-called Pasteur's quadrant of "use-inspired basic research" (Stokes 1997, p. 73ff). For the presented project, this means that the concrete product of the project—the teaching–learning arrangement—can now be integrated into a textbook and serves as material for in-service training. For the German curriculum in which probability usually starts only in grade 8 with Laplace definitions, the project explores a curricular innovation to start with phenomena of random (i.e. the empirical law of large numbers) already in grade 6 or 7. By this earlier start via data from chance experiments, we intend to contribute to the learning goal of context-adequate choice of stochastic conceptions.

At the same time, the learning arrangement offers the context for basic research that produces fundamental insights into the dynamics of stochastic learning processes and allows substantial contributions to theory development (see below).

4.2 Focus on Processes of Learning Rather than Only on Momentary States of Learning

Whereas the majority of empirical research studies in stochastics focus on momentary *states* of learning, our project aims at understanding the *dynamics* of the stochastic learning *processes* in terms of horizontal conceptual change processes. For this focus, we developed a theoretical conceptualisation of conceptions as webbings of constructs that consist not only of propositions, but also of the construct-elements function, situational context, and stochastic context. By this conceptualisation, we could describe the learning pathway of two girls as emergences of distinctions of stochastic contexts. In other parts of the project, the conceptual change takes place along with the merge of situational contexts or functions that were first perceived as isolated.

4.3 Focus Also on the Learning Content, not Only Teaching Methods

It is typical for Didactical Design Research to gain insights not only into content-independent learning pathways, but to focus also on the "what-question", that van den Heuvel-Panhuizen (2005) emphasises as crucial for Didactics. Here concretely, we could specify the empirical law of large numbers (i.e. the experience of rather stable frequencies for large numbers of throws) as an important prerequisite for understanding stochastic concepts. More precisely, the distinction between short and long term contexts could be decomposed into the distinction of one versus many games (which is crucial for every general pattern) and into small versus large totals of throws (which is typical for random phenomena). By these decompositions, we could elaborate the local teaching–learning theory with respect to the learning contents.

4.4 Mutual Interplay of Design and Research in All Iterative Cycles

Iterative cycles of design and research have often been emphasised as a core element of the methodology of Design Research (e.g. Gravemeijer and Cobb 2006). These cycles take place in design experiments that allow a versatile and flexible approach on teaching and learning and take into account the complexity of processes. In the presented project, we conducted five cycles and experienced the necessity to relate the results of each cycle with each other. The interplay is not only given by iterativity, but basically by seriously drawing consequences from one part of work for the other work. For example, the empirical insights change the theoretical perspective on the content and the design.

Of course, the project is far from being finalised. Instead, new interesting questions have emerged: In terms of *research*, it would, for example, be interesting how students are able to activate their gained knowledge on the distinction of short term and long term perspectives in new situations and in how far they are able to build new knowledge upon this (such as the introduction of different interpretations for probabilities). The challenge for the *design* will be to support students in doing so by creating new teaching–learning arrangements.

References

Aspinwall, L., & Tarr, J. E. (2001). Middle school students' understanding of the role sample size plays in experimental probability. *The Journal of Mathematical Behavior*, *20*, 229–245.

Barab, S., & Squire, K. (Eds.) (2004). Special issue: design-based research: clarifying the terms. *Journal of the Learning Sciences*, *13*(1).

Brown, A. L. (1992). Design experiments: theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences*, *2*(2), 141–178.

Burkhardt, H. (2006). From design research to large-scale impact. Engineering research in education. In J. van den Akker, K. Gravemeijer, S. McKenney, & N. Nieveen (Eds.), *Educational design research* (pp. 121–150). London: Routledge.

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

Confrey, J. (1990). A review of the research on student conceptions in mathematics, science and programming. In C. Cazden (Ed.), *Review of research in education* (Vol. 16, pp. 3–56).

diSessa, A. (1993). Towards an epistemology of physics. *Cognition and Instruction*, *10*(2–3), 105–225.

Duit, R., & Treagust, D. F. (2003). Conceptual change: a powerful framework for improving science teaching and learning. *International Journal of Science Education*, *25*(6), 671–688.

Fischbein, E. (1982). Intuition and proof. *For the Learning of Mathematics*, *3*(2), 9–19.

Gravemeijer, K., & Cobb, P. (2006). Design research from a learning design perspective. In J. Van den Akker, K. Gravemeijer, S. McKenney, & N. Nieveen (Eds.), *Educational design research* (pp. 17–51). London: Routledge.

Gropengießer, H. (2001). *Didaktische Rekonstruktion des Sehens. Wissenschaftliche Theorien und die Sicht der Schüler in der Perspektive der Vermittlung*. Oldenburg: DIZ.

Hußmann, S., & Prediger, S. (2009). Je größer die Wurfanzahl, desto sicherer die Wette: Mit dem Spiel Wettkönig den Zufall auf lange Sicht erkunden [The larger the throw total, the more secure is the bet: Exploring chance in the long term with the game Betting King]. *PM. Praxis der Mathematik in der Schule*, *51*(25), 24–29.

Hußmann, S., Leuders, T., Prediger, S., & Barzel, B. (2011). *Kontexte für sinnstiftendes Mathematiklernen (KOSIMA)—ein fachdidaktisches Forschungs- und Entwicklungsprojekt. Beiträge zum Mathematikunterricht 2011* (pp. 419–422). Münster: WTM Verlag.

Kattmann, U., & Gropengießer, H. (1996). Modellierung der didaktischen Rekonstruktion. In R. Duit & C. v. Rhöneck (Eds.), *Lernen in den Naturwissenschaften* (pp. 180–204). Kiel: IPN.

Komorek, M., & Duit, R. (2004). The teaching experiment as a powerful method to develop and evaluate teaching and learning sequences in the domain of non-linear systems. *International Journal of Science Education*, *26*(5), 619–633.

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

Moore, D. S. (1990). Uncertainty. In L. S. Steen (Ed.), *On the shoulders of giants: new approaches to numeracy* (pp. 95–137). Washington: National Academy Press.

Plaga, V. (2008). *Ein Spiel des Zufalls—Analyse von Erkundungen stochastischer Prinzipien durch Grundschüler*. Master equivalent thesis, University of Dortmund. Supervised by S. Prediger.

Posner, G., Strike, K., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: toward a theory of conceptual change. *Science Education*, *66*(2), 211–227.

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

Prediger, S. (2005). Auch will ich Lernprozesse beobachten, um besser Mathematik zu verstehen. Didaktische Rekonstruktion als mathematikdidaktischer Forschungsansatz zur Restrukturierung von Mathematik. *Mathematica Didactica*, *28*(2), 23–47.

Prediger, S. (2008). Do you want me to do it with probability or with my normal thinking? Horizontal and vertical views on the formation of stochastic conceptions. *International Electronic Journal of Mathematics Education*, *3*(3), 126–154.

Prediger, S., & Link, M. (2012). Fachdidaktische Entwicklungsforschung—Ein lernprozessfokussierendes Forschungsprogramm mit Verschränkung fachdidaktischer Arbeitsbereiche. In H. Bayrhuber, U. Harms, B. Muszynski, B. Ralle, M. Rotangel, L.-H. Schön, H. J. Vollmer, & H.-G. Weigand (Eds.), *Formate Fachdidaktischer Forschung. Empirische Projekte–historische Analysen–theoretische Grundlegungen* (pp. 29–46). Münster: Waxmann.

Prediger, S., & Rolka, K. (2009). Using betting games for initiating conceptual change. *Asian Journal of Educational Research and Synergy*, *1*(1), 61–71.

Prediger, S., & Hußmann, S. (2014, in press). Spielen, Wetten, Voraussagen—Dem Zufall auf die Spur kommen. In T. Leuders, S. Prediger, B. Barzel, & S. Hußmann (Eds.), *Mathewerkstatt 7*, Berlin: Cornelsen. (Textbook for German middle schools.)

Ron, G., Dreyfus, T., & Hershkowitz, R. (2010). Partially correct constructs illuminate students' inconsistent answers. *Educational Studies in Mathematics*, *75*(1), 65–87.

Schnell, S. (2013). Muster und Variabilität entdecken. *Konstruktionsprozesse kontextspezifischer Vorstellungen zum Phänomen Zufall*. Heidelberg: Springer Spektrum. PhD Thesis, TU Dortmund University (Supervisor: S. Prediger).

Schnell, S., & Prediger, S. (2012). From "everything changes" to "for high numbers, it changes just a bit" theoretical notions for a microanalysis of conceptual change processes in stochastic contexts. *ZDM—The International Journal on Mathematics Education*, *44*(7), 825–840.

Schoenfeld, A. H. (2006). Design experiments. In P. B. Elmore, G. Camilli, & J. Green (Eds.), *Handbook of complementary methods in education research* (pp. 193–206). Washington & Mahwah: American Educational Research Association and Erlbaum.

Schwarz, B., Dreyfus, T., & Hershkowitz, R. (2009). The nested epistemic actions model for abstraction in context. In B. Schwarz, T. Dreyfus, & R. Hershkowitz (Eds.), *Transformation of knowledge through classroom interaction* (pp. 11–41). London: Routledge.

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.

Stokes, D. E. (1997). *Pasteurs quadrant*. Washington: Brookings Institution Press.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, *76*(2), 105–110.

Van den Akker, J., Gravemeijer, K., McKenney, S., & Nieveen, S. (Eds.) (2006). *Educational design research*. London: Routledge.

Van den Heuvel-Panhuizen, M. (2005). Can scientific research answer the 'what' question of mathematics education? *Cambridge Journal of Education*, *35*(1), 35–53.

Wittmann, E. Ch. (1995). Mathematics education as a 'design science'. *Educational Studies in Mathematics*, *29*(4), 355–374.

Counting as a Foundation for Learning to Reason About Probability

Carolyn A. Maher and Anoop Ahluwalia

Abstract Based on findings from long-term and cross-sectional studies in a variety of contexts and across a variety of ages, we have found that in the activity of problem solving on strands of counting and probability tasks, students exhibit unique and rich representations of counting heuristics as they work to make sense of the requirements of the tasks. Through the process of sense making and providing justifications for their solutions to the problems, students' representations of the counting schemes become increasingly more sophisticated and show understanding of basic combinatorial and probabilistic reasoning.

1 Introduction

The Roman philosopher, Cicero, remarked: "Probability is the very guide of life." In today's modern world where informed citizens need to understand the language of statistics and probability to absorb the information presented in news, media, health reports, sales and advertising, Cicero's vision still applies. In 2000, the National Council of Teachers of Mathematics (NCTM) placed increased emphasis on probability and statistics learning in the K-12 curriculum by including these subjects as one of the five major content standards and recommending these as major content areas. The NCTM indicated that learning to reason probabilistically is not knowledge that learners intuitively develop. They point to the importance of including these subjects in the mainstream curriculum. Specifically, they indicate: "The kind of reasoning used in probability and statistics is not always intuitive, and so students will not necessarily develop it if it is not included in the curriculum" (NCTM 2000, p. 48).

Several studies have shown that a strong foundation in counting methods is an important prerequisite for fundamental probabilistic reasoning (Alston and Maher

C.A. Maher (✉)
Rutgers University, 10 Seminary Place, Room 231, New Brunswick, NJ 08901, USA
e-mail: carolyn.maher@gse.rutgers.edu

A. Ahluwalia
Brookdale Community College, 765 Newman Springs Road, Lincroft, NJ 07738, USA

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_30,

2003; Benko 2006; Benko and Maher 2006; Kiczek et al. 2001; Francisco and Maher 2005; Maher 1998; Maher and Muter 2010; Maher et al. 2010; Shay 2008). The Rutgers Longitudinal Study, initiated in 1987, engaged students in working on a strand of counting and probability tasks in which the students were challenged to build and justify their solutions to the problems (Maher et al. 2010). The research has shown that under particular conditions in which students are invited to collaborate, share, revise and revisit ideas, sophisticated counting schemes are built (Maher 2005; Maher and Martino 1997). The counting schemes are represented in a variety of ways and, over time, are elaborated and expressed in more abstract forms, becoming increasingly more sophisticated and formal (Davis and Maher 1997; Maher and Weber 2010). As the students recognize connections between and among representations in their solutions to counting problems, they identify isomorphisms between problem tasks. Their collection of representations provides a rich resource for building even more complex schemes and for solving more difficult problems (Maher 2010; Davis and Maher 1990; Maher and Martino 2000).

A basic foundation for successful reasoning in solving probability problems is the recognition of an adequate representation of the sample space for an experiment. In this chapter, we give several examples of students' use of representations in building, refining and extending the sample space for counting problems. We show how the richness of the students' early counting schemes provides a strong foundation on which to build probabilistic ideas over time. We include data from students' early problem solving to later years. Finally, in episodes that we provide from our longitudinal study, we show how increasingly more complex counting schemes and representations are built in search of understanding the properties of Pascal's Pyramid (Ahluwalia 2011).

2 Theoretical Perspective

When making mathematical decisions in a novel situation, the learner often needs to build and rebuild more efficient representations and make them flexible to deal with the requirements of the new task (Pirie et al. 1996). Even with very basic situations, probabilistic reasoning can be complex and subtle and the learner needs to rely on models and representational strategies that can be extended easily to support the growth of probabilistic ideas (Maher and Speiser 2002). The development of probabilistic thinking requires reflective building over time (Maher and Speiser 2002). The notion of reflection in learning has been a core principle for our research. This perspective provides students with opportunities to build fundamental ideas in counting and to revisit, extend, and modify solutions over time. To support reflective learning, the researchers often deferred closure of solutions so that students could build convincing arguments and support their solutions to the problems posed.

Davis and Maher (1990) indicate that to think about a mathematical situation, a student cycles through a process of building, validating and connecting representations of the data to the representations of the previously acquired knowledge. Once the student has created a representation for the data, the student searches for representation of relevant knowledge in his/her personal inventory of mental representations that can be used to solve the problem. If these knowledge representations exist and make sense, the student then connects the data representations to the knowledge representations. If the student is not able to find relevant knowledge representations in the mental library, then he/she might need to rethink the representations or even build new knowledge representations. This is not an effortless step for the student, as the student might have to modify, reject or extend an original knowledge representation to successfully support the new findings or justify the arguments. This step is further complicated by the common misconceptions about probability as an outlier in the data set might hinder this process of building and validating the representational connections (Shay 2008).

3 Background of the Longitudinal Study

The longitudinal study began as a partnership between the Kenilworth Public Schools and Rutgers University. It originated at the Harding Elementary School in a blue-collar neighborhood of New Jersey. Led by Carolyn Maher, the project began with a group of graduate students who were the original Rutgers Team. They were invited to provide professional development to the teachers of the school. This led to the longitudinal study of the development of mathematical ideas and ways of reasoning in students. The work was supported first by the Kenilworth Public Schools, several foundation grants and a series of National Science Foundation grants.[1] The research has been ongoing for a quarter of a century and has produced a collection of videos that are being prepared for storage in a Repository, the Video Mosaic Collaborative (VMC), at Rutgers. The work with students is described in detail elsewhere, with the counting strand as the initial content (Maher et al. 2010). The Rutgers team visited classrooms to work with the students several times each year. The approach was to pose interesting tasks for the students to work together to solve. The students were invited to collaborate, share and discuss their ideas, and to present their findings. They were asked to justify their solutions to each other and to the researchers (Maher and Uptegrove 2010). The students were provided with open-ended tasks and extended time to develop their ideas and revisit these ideas after a long period of time. The design was for minimal intervention from the researchers. When disagreements occurred, they were left unresolved until students were ready to revisit and support their arguments. The expectation was for the students to work collaboratively and share the representations for their ideas. This required considering and evaluating each other's ideas. In so doing, they recognized equivalent representations and often adapted another's ideas if judged to be useful. Over the years,

[1]Grants: NSF awards: MDR-9053597, directed by Robert B. Davis and Carolyn A. Maher. REC-9814846, REC-0309062 and DRL-0723475, directed by Carolyn A. Maher.

participants of the longitudinal study were invited to work on tasks from several different mathematical strands: counting, probability, algebra, geometry, pre-calculus, and calculus. As a part of the counting and probability strand, students worked on tasks to find the sample space of experiments, discussed fairness of the design of dice games, and found and justified the probability of events such as in the World Series Problem.

The students participating in the study continued to meet with the researchers voluntarily through their high-school years and beyond. High-school sessions that were conducted after school involved small-group problem solving sessions and individual and group interviews. A two-week institute was held the summer after the students' junior year and some earlier participants who had moved or changed to private/parochial schools joined the summer sessions. All of the activities were video taped and transcribed to produce a unique collection of video and related metadata consisting of students' written work and problem tasks as well as detailed analyses produced from dissertation work.

4 Methodology

For the episodes discussed in this chapter, we have several sources of data: video of the session, transcripts of the session, along with the original student handwork and the researcher's field notes. To analyze this data, we first watched the video data to identify points of interest related to the creation of representations and building of combinatorial ideas along with emergence of probabilistic arguments. These events were our characterized as critical events (Powell et al. 2003). Then we created a timeline of these critical events to build a story line and provided a detailed narration of the events. To provide triangulation of data we validated our findings using the student work and researcher's field notes wherever available. We also used the detailed transcripts that were available to closely study and analyze the emergence and elaboration of arguments related to counting and probabilistic ideas. We provide excerpts from the student dialog, their handwork and video snapshots to illustrate the findings.

5 Results

Probabilistic reasoning is complex as the models built by students to understand the data can be non-representative and reasoning from small samples can contribute to extending the distortions (Maher 1998). Also, the data collected by students and the model created to interpret phenomena can mislead students in their statistical predictions (Maher 1998). Providing students with ample time and opportunities to revisit their problem solving can help in the development of a more refined understanding of statistics and probability ideas. We share here several episodes from the Rutgers longitudinal study that illustrate how students worked to make sense of the meaning of some combinatorial tasks. They expressed their solutions with a variety of creative representations of sample spaces for their experiments. Consider, for example, the problem of determining the fairness of games.

5.1 Is the Game Fair?

A group of researchers from Brazil, Israel, and the United States carried out a cross-cultural investigation concerning the probabilistic and statistical reasoning in children and adult learners (Amit 1998; Fainguelernt and Frant 1998; Maher 1998; Speiser and Walter 1998; Vidakovic et al. 1998). The studies investigated in detail the ideas built by students about dice games as students in the three countries worked on the same tasks, across several grade levels—elementary to college level. For example, all students were invited to work on the same two problem tasks to determine the fairness of games.

Game 1 (A game for two players) Roll one die. If the die lands on 1, 2, 3, or 4, Player A gets 1 point (and Player B gets 0). If the die lands on 5 or 6, Player B gets 1 point (and Player A gets 0). Continue rolling the die. The first player to get 10 points is the winner. Is this a fair game? Why or why not?

Game 2 (Another game for two players) Roll two dice. If the sum of the two is 2, 3, 4, 10, 11, or 12, Player A gets 1 point (and Player B gets 0). If the sum is 5, 6, 7, 8, or 9, Player B gets 1 point (and Player A gets 0). Continue rolling the dice. The first player to get 10 points is the winner. Is this a fair game? Why or why not?

In the United States, Maher and colleagues challenged fifth-grade students to play the games. The students concluded, almost instantly, that the first game was unfair. For the second game, students discussed several ways to modify the game so that each player would have the same number of chances to score a point. Some students claimed that Player A had a clear advantage since it had more sums to score by than Player B. Some students indicated that certain sums were easier to roll and consequently, the game was probably fair as these easier sums made up for the fact that Player B had one less sum to score by. A student, Jeff, explained to his classmate, Romina, that "snake eyes" (sum of 2) and "boxcars" (sum of 12) were the most difficult sums to roll and that 7 was the easiest. The students worked on this activity in groups and played the game several times. As the students recorded scores for Players A and B, they began to reevaluate their initial hypothesis. As the session ended, students were encouraged to play the game at home and report back during the next session on their findings and explanations.

In the second session, the students again collaborated to find a way to make the game fair. Jeff produced a chart to record what he considered were the possible outcomes for a particular sum from 2 to 12. He listed 21 elements for his sample space where 13 of the sums he listed favored Player B and 8 favored Player A. He concluded that since 21 is odd, there is no fair division of events possible. Another

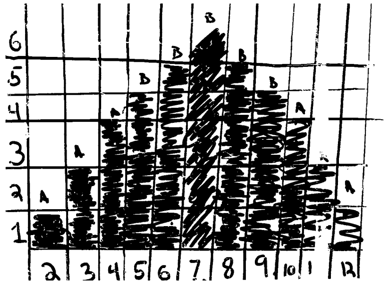


Fig. 1 Stephanie's histogram for the outcomes of rolling a pair of dice

student Stephanie, along with Ankur and Milin, worked with a different model that had 36 elements. Stephanie created a chart to represent her possibilities that included the colored histogram shown above (Fig. 1).

At this point in the sessions, both Jeff and Stephanie appear to view the points for their sample spaces as equally probable. The students went on to have a discussion with the teacher that focused on whether the order of events matter in counting the sample space, that is, whether an outcome of $(1, 2)$ is different than an outcome of $(2, 1)$ given that they both result in a sum of 3. A brief excerpt of the conversation that followed is shared below.

Teacher: Ok. So Amy is telling me there is one way you can get three. And what's that way?

Amy: 2 and 1.

Teacher: Amy says you can get three by 2 and 1 ... one way... and she can get six by?

Jeff: 2 and 4, 3 and 3, 5 and 1.

Teacher: Three ways. Do you all agree with that?

Ankur: No.

Teacher: Ankur, Ankur doesn't agree with that.

Ankur: I say for three, there's 2 and 1 and 1 and 2, because 2 is on one die and 2 is on the other die and 1 is on the one die and 1 is also on the other die. You can have 2 on this die and 1 on this die... or you can have 1 on this die, whatever it is, 2 on this die.

During this conversation, Stephanie was standing by the projector with her graph of outcomes. While another student Michelle tried to build on what Ankur was saying, Stephanie moved to a table and shared her graph with Jeff. Jeff then interrupted Michelle to say that what Ankur said actually made sense and decided to discard his earlier model in which he had listed 21 outcomes.

Jeff: And that makes it two chances to hit that even though it's the same number. It's two separate things on two different dice.

Stephanie: Therefore, there's more of a chance. Therefore, there are two different ways. Therefore, there are two ways to get 3.

Jeff: And that throws a monkey wrench... and that just screws up everything we just sort of worked on for about the past hour.

This conversation satisfied Stephanie and she went on to explain the outcomes in her model. Together, through a group discussion, the students correctly found the sample space for rolling a pair of dice and agreed upon the 36 outcomes. The sharing of ideas by students gave them an opportunity to reconsider their initial interpretation of the outcomes of the game, enabling them to think together and explore alternative ideas proposed by peers. For this group, it led to a correct listing of all possible outcomes of the game, enabling the creation of a correct sample space.

5.2 Tetrahedral Dice

In the seventh grade, after working with sample space for three 6-sided dice, the students were introduced to tetrahedral dice so that they could collect data that was manageable to verify their conjectures about the sample size. Dann et al. (1995) outlined three episodes from these sessions where students attempted to list all possibilities when rolling three tetrahedral dice. The researcher first asked the students to consider rolling only two dice. After some discussion, the students reported that

Fig. 2 Ankur's sample space for tetrahedral dice

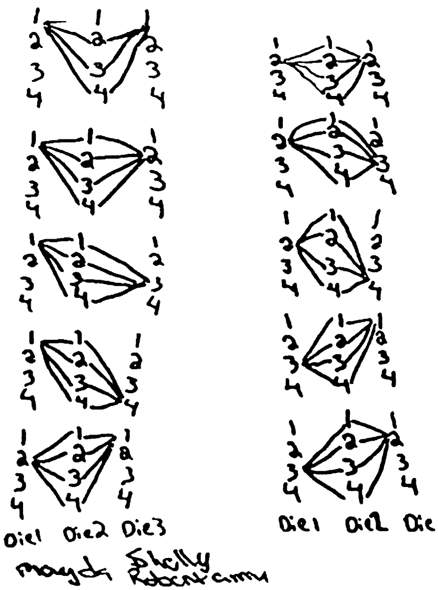


Fig. 3 Michelle's first tree diagram for the tetrahedral dice

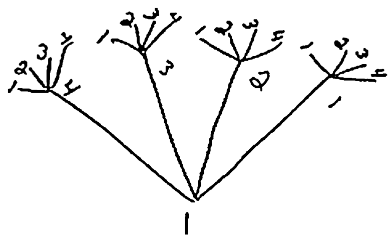


Fig. 4 Michelle's second tree diagram for the tetrahedral dice

there are 4 outcomes for one die and 16 outcomes for two dice. When the teacher asked Ankur to show a way of representing all of the outcomes, he made the following figure that connected the two dice by drawing lines from 1 on the first die to the 1, 2, 3, and 4 on the second die and then repeating this pattern for 2, 3, and 4 on the first die. As a result he drew 16 connections as shown in the figure below. When the teacher asked him to explain his figure, Ankur created a list of ordered pair as shown below (see Fig. 2).

The teacher then encouraged Ankur to explain how many outcomes would be possible when rolling three dice. Ankur's first response was that there would be 64 outcomes. When the teacher asked Ankur to show how the 64 outcomes are generated, Ankur generalized his previous findings and proposed to add four more numbers to his figure, the third column (on left in Fig. 2). In the remaining episode, the students created several other tree diagrams to represent their counting schemes. On the following day the students were asked to explain their written work. Another student, Michelle shared the tree diagram that she created as shown below (see right of Fig. 3).

Michelle explained that each column of numbers represented a die and the connecting lines represented the outcomes that were possible when tossing three dice at the same time. Responding to the researcher's request to show all 64 combinations that the group had proposed as a solution on the previous day, Michelle expressed uncertainty that her figure captured all of the 64 outcomes and began referring to another tree diagram that she had created as shown in Fig. 4.

Michelle recollected that some time ago, they had worked on a towers activity where the students had created a "tree thing". She explained that the first tier of the figure above represents "what you would roll on the first die" and the second tier was what you would roll on the second die and similarly the third tier were the outcomes from the third die. Michelle had used an earlier representation for the towers activity and adapted it to represent the outcomes for the tetrahedral dice to optimally represent the sample space.

As the session came to an end, Ankur suggested a generalization of the findings and proposed a rule indicating that the number of possibilities can be determined by the number of sides of the die raised to the power equaling the number of the dice that are being tossed. As such, he explained that for 4 tetrahedral dice, there would be 4^4 or 256 possibilities. Again, the students were able to create and modify their own representations to find the sample space for the tetrahedral dice and successfully created and revised their novel tree diagrams.

In addition, responding to the researcher's request to justify their findings resulted in students folding back to an earlier representation as well as creating and modifying a new one to show the details of the outcomes of the game. It is interesting too that a rule emerged that enabled the students to predict the correct number of outcomes without detailing the specifics of the events.

Now, we examine another set of examples in which students created and used tree diagrams to list possibilities of an experiment and find solutions to increasingly complicated combinatorial tasks.

5.3 More Tree Diagrams

The tree diagram is a widely used organizer for listing possibilities of a sample space that is traditionally introduced to students in an introductory course on probability. The previously discussed episode sheds light on how students in the longitudinal study decided to use the idea of a tree diagram on their own to verify the number of outcomes for tetrahedral dice. Sran (2010) traced work of a student Milin in detail over his elementary grades and discovered that Milin was the first student to discover and explain an inductive argument to his peers while building towers of a certain height using two-colored blocks. Sran (2010) outlined in great detail how Milin's inductive argument facilitated the creation of tree diagrams that was later adapted by several of his classmates and we recapture the findings here briefly.

In Grade 4, Milin and Michael worked on building five-tall towers selecting from red and yellow cubes. Milin and Michael initially created towers by trial and error and then found opposites for the towers that they had created. Milin also used a "cousin" strategy where he inverted the tower vertically to create a new tower. The students checked their work by looking for duplicates and found all 32 five-tall towers. Later, the researchers to discuss Milin's justifications for the 32 five-tall towers that he created, interviewed him. During the interview, Milin began to consider simpler cases and pointed out that there are four towers that can be built that are two-cubes tall and two towers that are one-cube tall when choosing from two colors. Milin later took the cubes home and worked on the activity further to observe the results for the simpler cases (Sran 2010).

After a couple of weeks, Milin reported in another interview that there were 16 four-tall towers and recorded the number of combinations that were possible for one, two and three-tall towers. Later in a third interview, Milin showed that the taller towers could be constructed from the shorter ones. He explained that to build four two-tall towers, one could start with the two one-tall towers (a blue and a black cube) and place either a black or a blue cube on each of the cubes and create four different two-tall towers (black–black, black–blue, blue–black and blue–blue). Eventually, Milin suggested that his rule for generating larger towers from the smaller towers would work for towers taller than five cubes and that there were 64 possible combinations of six-tall towers.

After three weeks, in a small-group interview, a simpler version of the problem was chosen to help students share their strategies to account for all possible three-tall towers using two colors (Maher and Martino 1996a, 1996b). Although the activity was to find all three-tall towers, the researcher asked students how many six-tall towers can be built selecting from two colors and Milin answered “probably 64”. When he was asked to explain how he knew, Milin described his inductive argument, and suggested to multiply the previous answer by two (Sran 2010). Milin explained that there were two, one-tall towers; four, two-tall towers; and eight, three-tall towers, and the number of towers doubled for the next height as a cube of each of the two colors could be added on top of the two copies of the shorter tower. To explain his doubling rule, Milin drew a sketch of two, three-tall towers resulting from a two-tall tower by placing a red cube followed by a blue cube on its top.

In the same session, another student Stephanie presented an argument by cases to build all possible towers of a certain height (details in Maher and Martino 1996a, 1996b). Stephanie defended her argument with cases in the session but later indicated that she understood Milin’s inductive argument as well. As the session came to a close, several students used Milin’s argument to defend their solutions to the towers problem (Sran 2010; Maher and Martino 1997, 2000; Maher 1998). The inductive argument introduced by Milin initiated the idea of a tree diagram for the group and helped students discover how it can be used to create an exhaustive list of the outcomes of the experiment.

Later in fifth grade, Stephanie demonstrated Milin’s inductive idea by creating a tree diagram using Unifix cubes and presented it to a group of students that included Robert. Ahluwalia (2011) studied Robert’s work in great detail over a 16-year period and later conducted task-based interviews with him where he was invited to explore the properties of the Pascal’s Pyramid. Ahluwalia (2011) outlined how Robert reused tree diagrams in his adulthood in sophisticated manners to solve problems related to the Pascal’s Pyramid and we share the results briefly here.

In fifth grade, Robert observed Stephanie explain to the group in her own words how the “parent” towers gave rise to “children” towers. Stephanie repeated the strategy that Milin had explained by creating two copies of each tower one-tall and placed two red (dark) blocks and two yellow (light) blocks in the second row of her tree diagram. On top of the two red blocks, she added one red and one yellow block to create two “children” for the “parent” red tower as red–red and red–yellow. Similarly, she duplicated the yellow parent tower and created the yellow–red and yellow–yellow “children”. She explained that for the parents that were one-tall, only four children two-tall were possible. She repeated the process to create three-tall and four-tall towers as shown in Fig. 5. Robert closely listened to Stephanie’s presentation and agreed with her justifications.

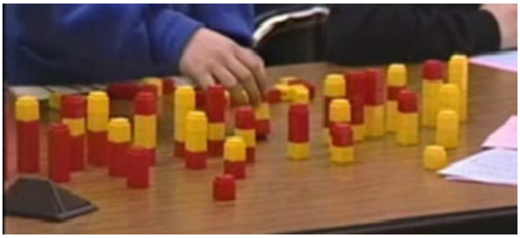


Fig. 5 Stephanie explains her tree diagram for building towers

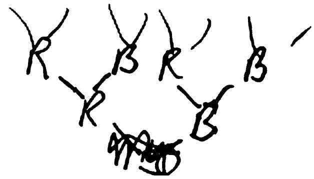


Fig. 6 Robert's sketch of a tree diagram

Later in grade 11, Robert worked on an extension of the towers activity. He used the tree diagram as a sketch to justify to the researchers that he had found all towers possible using two colors (Fig. 6). He explained using his sketch that the one-tall towers only had two options, R for red and B for blue. From each of the one-tall towers, he drew a set of two branches coming out and placed an R and a B on each set of branches. He indicated that the two sets of branches together gave rise to four two-tall towers that he listed as RR, RB, BR, and BB. He argued that there were no other towers possible for the two-tall case, as there was nothing else to add to the one-tall towers. He continued this inductive argument to justify that he had found all possible 16 four-tall towers. In this episode, Robert created a symbolic representation for the tree diagram (sketch) to recapture what Stephanie had shared with him using the Unifix cubes. Robert also demonstrated through his justifications that he had understood and assimilated Milin's inductive argument as reiterated by Stephanie to his group in the fifth grade. This is a powerful example of how ideas traveled among the students and how the individual students personally adapted the ideas that were collaboratively built.

To share a more sophisticated example of Robert's use of the tree diagram, we share results from his third post-graduate interview with Ahluwalia and Maher where he used the Zome tools (connecting plastic rods and spheres) to build a 3-D model of the Pascal's Pyramid. Pascal's Pyramid is an intricate mathematical identity that lists the coefficients resulting form the expansions of the trinomial $(a+b+c)^n$. Robert had previously interviewed twice with Ahluwalia where he had constructed sketches to illustrate cross-sections of the Pyramid and also built an all-blue model with Zome tools for the Pyramid. In this episode, Robert chose two-colored rods to create his model and explained that the two colors represented two operations in the Pascal's Pyramid. He explained that the blue rods were used to represent addition between the terms of a layer and the red rods were used to represent multiplication by the three variables a, b, and c as one moves between any two consecutive layers.

Marjory, a graduate student, was an observer in this episode, and Robert was helping her discover the connections between the activity of building towers using three-colored Unifix cubes and the structure of Pascal's Pyramid. The researchers encouraged Robert to explain to Marjory how the second layer of the Pyramid was derived from the first layer. At this point, Robert chose to use the Unifix cubes and created an interesting tree diagram. He started his tree with three cubes, one each of the yellow, blue and red color. Next, he connected these cubes with blue rods to represent addition between the terms of the layer $(a+b+c)^1$, in which the three cubes and the blue rods represented the sum, $a+b+c$. He then placed three sets of

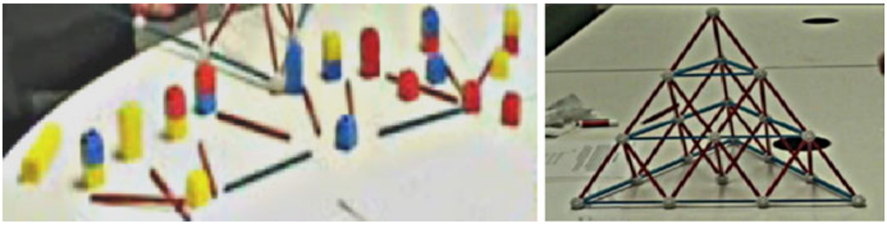


Fig. 7 Robert's Unifix tree diagram (*left*) and two-colored model of the Pyramid using Zome tools (*right*)

three red-colored rods coming out of each of the one-tall towers. He explained that the three towers of a color would be multiplied by each of the three variables, a, b, and c. He placed a yellow, a blue, and a red cube next to one set of three red rods to represent his multiplication as shown in a sketch of his model (left of Fig. 7).

Next, Robert created the two-tall towers that resulted from the multiplication. For instance, when the term represented by the red cube was multiplied by the term represented by the yellow cube, it gave rise to the red–yellow tower. He continued to generate his nine two-tall towers resulting from the nine products.

The researchers pointed out to Robert that there were only six spheres in the second layer of the Pyramid, rather than nine. Robert responded that some of the two-tall towers represented like terms that needed to be combined. He explained that the red–yellow and yellow–red towers should be grouped together to represent two red/yellow towers. He further explained that the towers with two different colors were equivalent to the $2ab$, $2bc$, and the $2ac$ terms that result from the expansion of $(a+b+c)^2$ and that the three variables a, b, and c be represented by the three colors, yellow, blue, and red.

This instance demonstrates an impressive accomplishment of Robert, that is, the fluent building of an isomorphism among three mathematical tasks: building towers selecting from three colors, expansion of trinomials, and the Pascal's Pyramid. Making use of his understanding of the inductive argument that Milin had introduced years earlier and the tree-diagram construction of towers that Stephanie had shared with him, Robert had the basic ideas needed to create a foundation for his later understanding of the structure of Pascal's Pyramid. The tree diagram used by Robert in this episode exemplifies the resilience and fluency of the meaningful representations that were created earlier with his classmates. It also demonstrates how these representations can be extended to accommodate increasingly sophisticated mathematical relationships and demonstrate intricate isomorphisms.

5.4 Ankur's Challenge

Another interesting combinatorial problem that has been documented in the longitudinal study was posed to a group of students by a tenth-grader, Ankur. The problem came to be known as Ankur's Challenge:

> Find all possible towers that are four cubes tall, selecting from cubes available in three different colors, so that the resulting towers contain at least one of each color. Convince us that you have found them all (Maher 2002; Maher et al. 2010).

We share an episode from Ahluwalia (2011) in which Robert connected his solution to Ankur's Challenge to a component of Pascal's Pyramid. In an informal conversation with mathematician, Professor Todd Lee from Elon University, Robert attempted to sketch layers of the Pascal's Pyramid using markers on the white board. The fourth layer of the Pyramid has terms resulting from expansion of $(a+b+c)^4$ and is a rather challenging layer to list mentally. While working on the fourth layer, Robert was unclear what terms would be included in the middle of this layer. Researcher Maher was observing the conversation and Robert's work. She invited Robert to think about how the solution to Ankur's Challenge could be represented on the Pyramid. Robert first pointed out that the solution might be located in the third layer of the Pyramid and then quickly noted that since the third layer corresponded to towers three-tall, a fourth layer would be required to illustrate four-tall towers. He conjectured that the towers for Ankur's Challenge would lie in the middle of the fourth layer and that there would be three 12s in middle of the layer. At this point, Robert did not yet offer an explanation how the terms from the third layer gave rise to the 12s in the middle of the fourth layer. He had created several lists of the terms from the expansion of $(a+b+c)^3$ and was investigating how a multiplication by the three variables a, b, and c would yield terms for $(a+b+c)^4$.

As Robert continued to count (and sometimes miscount) the terms in the fourth layer, he decided to sketch the towers on a white board to represent the terms in the middle of the fourth layer. He created sketches of the towers and realized that the terms in middle of the Pyramid's fourth layer consisted of $12a^2bc$, $12ab^2c$, and $12abc^2$. The act of sketching out the towers and associating the three variables of the terms to the three colors of the towers aided Robert in addressing his confusion about the terms that were alike or not in the fourth layer (Ahluwalia 2011). He then quickly pointed out that the Pascal's Pyramid on its three planar surfaces was a Pascal's Triangle and only involved two of the three variables at a time, concluding that only the middle of the Pyramid consisted of terms where all the three variables, a, b, and c were present. He then explained that the solution to Ankur's challenge was given by the three 12s in middle of the fourth layer that together give 36 towers that are four tall and have at least one block of each color. Robert illustrated that the term $12a^2bc$ represented a tower with two blocks of the color "a", a block of color "b" and a block of color "c". Similarly, Robert contended that the terms $12ab^2c$ and $12abc^2$ were four-tall towers that had a block each of the three colors yielding a total of 36 towers that satisfy Ankur's Challenge.

This was a rather unique representation of the solution to Ankur's Challenge, and Robert had once again created a bridge between several mathematical representations of the solutions to a problem. He illustrated an isomorphism between the building towers activity, expansions of the trinomials, and the structure of the Pascal's Pyramid by giving very concrete explanations for how towers with specific properties related to specific parts of the Pyramid and exquisitely solved Ankur's Challenge.

5.5 *World Series Problem*

In the eleventh grade (1999), students were introduced to the following problem called the "World Series Problem":

> In a World Series, two teams play each other in at least four and at most seven games. The first team to win four games is the winner of the World Series. Assuming that the teams are equally matched, what is the probability that a World Series will be won: (a) in four games; (b) in five games; (c) in six games; and (d) in seven games?

Kiczek (2000) analyzed four separate sessions in detail where students discussed solutions to this problem and we share highlights of the first session here. The group of students who were initially invited to work on this problem, in an after-school session, included Ankur, Brian, Romina, Jeff, and Michael. Immediately, the students agreed that there are only two ways for the series to end in four games. Romina offered the two cases as AAAA and BBBB to represent that either team A wins all four games or team B wins all four games. Jeff remarked that it would be the "hardest" to win the series in four games and everyone agreed. Brian related the series to flipping of a coin and students discussed if they should multiply one-half by itself to find the probability of the series ending in four, five, six, or seven games as for each game both the teams have a 50 % chance of winning. However, they soon realized that consecutively multiplying one-half with itself would lead to a smaller fraction indicating that it is less likely for a series to end in seven games compared to four games. This was counterintuitive to their initial guess that ending the series in the four games is the "hardest".

Ankur began to list the possibilities that team A would win the series and remarked that team A had to win the last game for the series to last the particular number of games (Fig. 8).

The group agreed with Ankur's list and that there are eight ways for the series to end in five games but they were unsure of the denominator for the probability. The students considered some of the earlier combinatorial tasks such as building towers and counting pizzas where they had discovered solutions of the form 2^n. Romina suggested that the denominator should be 2^5 because there are five possible games. Ankur explained that probability for the series to end in four games should be $2/2^4$ and Jeff agreed with him and proposed $8/2^5$ for the five game series. The students went back to discussions of "spaces", "slots", or "blanks" and used language of towers, pizzas, and binary lists to discuss the total number of possibilities and discussed that for each game (or "space") there were only two possibilities, namely the events that either team A or team B wins.

Jeff suggested that they add up all the probabilities and subtract from one using the idea of a complimentary event to find the probability for the last case, the seven-game series. Using the previously calculated probabilities for a four, five, and six-game series, Jeff calculated $13/32$ for the series to end in seven games. This answer did not agree with the solution of $10/32$ that Ankur had calculated for a seven-game series from his list. At this point, the students reviewed their work and found that for a part of the problem, Brian had missed two winning outcomes in the six-game

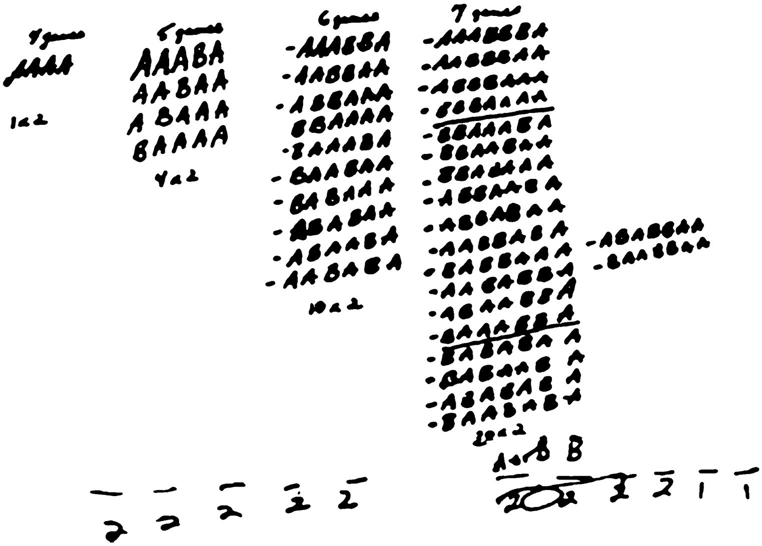
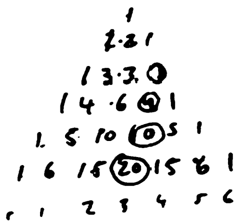


Fig. 8 Ankur's list for possible ways a World Series could end

Fig. 9 Michael's Pascal's Triangle



series and after revisions all their probabilities added up to 1. The students correctly concluded that the probability to win the series in four games is $1/8$, in five games is $1/4$, in six games is a $5/16$ and for seven games is a $5/16$ as well. Jeff was not sure that they had a "really concrete mathematical backing" for their solution. The teacher/researcher asked them if they could convince someone else of their solution and Michael shared what he had been working on by himself. He had listed first seven rows of the Pascal's Triangle. He pointed out that one of the diagonals contained the numbers: 1, 4, 10, and 20, and when these numbers are doubled, the total possibilities for the four-game, five-game, six-game, and seven-game series are found (Fig. 9).

Ankur conjectured that this connection to Pascal's Triangle could prove that they did not miss any cases. A month earlier, Michael had been able to explain the connection of the pizza toppings to the addition rule of Pascal's Triangle using his code of binary strings and he remarked that code of As and Bs for the World Series was similar to the codes of 1s and 0s. In about an hour, the students had correctly solved the World Series problem, explained their strategies, and justified the connection of their solution to the entries of Pascal's Triangle (Kiczek 2000).

Kiczek (2000) outlined another session a week later where the same group of students, except Brian, presented a slightly different justification for their solution to the problem and again correctly defended their answer. Later in this session, the students were introduced to an incorrect solution that a group of graduate students offered as their solution for the problem. The researcher indicated that she was confused about the discrepancy and asked the group to examine the work of the graduate students whose sample space consisted of 70 outcomes. Initially, this second solution was unsettling and some of the students questioned their own correct solution (Kiczek 2000). Michael, for example, indicated that there could be a total of 70 possibilities and that the series could end in four, five, six, or seven games. Jeff questioned why the probability of the series ending in six or seven games was the same in their correct solution. The graduate students' work presented different probabilities for the six and the seven-game series, and Jeff began to consider their solution as well. The researcher challenged the students to provide support for the solution; however, Michael was unable to find a flaw in either of the arguments and expressed confusion. Researcher Maher invited Robert to help Michael reach a conclusion about the solution and arranged for another session to discuss the dilemma (Kiczek and Maher 2001). The session is detailed in Ahluwalia (2011) and shared briefly here.

Michael had worked on the World Series problem prior to this session and Robert had not. Michael had two different solutions for the problem from the previous sessions as described above. After Robert expressed understanding of the statement of the World Series problem, Maher stepped out of the room. Michael began by using binary notation and represented 1 as one of the teams and 0 as the other team. For instance, he used 1111 to represent the event that "team 1" won first four games and therefore won the World Series. Michael also shared with Robert that they could list possibilities for team 1 winning and then multiply the number of possibilities by 2 to account for the event that team 0 would win. Robert began to list possibilities using the binary notation as well. Robert had previously worked with binary notation in his elementary grades when he worked on an earlier counting problem (Ahluwalia 2011).

Robert created a list of 10 possibilities for a six-game series in which the team 1 won. He then multiplied 10 by 2 and concluded that there are 20 ways for the series to end in six games as each team could win in 10 ways, and Michael worked on different computations. Robert then shared his list with Michael, who indicated that he remembered 20 as an answer from his previous experience with the problem.

On close inspection of Fig. 10, we see that Robert had a system to generate his list that involved controlling for a variable (Ahluwalia 2011). Robert started his list with 00 together in the first spot (001111) to represent that team 0 wins the first two games followed by team 1 winning the following four games. For the next entry, he inserted a 1 in between the 00 (010111) followed by inserting two 1s in between the 00 (011011) and so on. Robert was controlling the two-out-of-six games that team 0 could win to list all the ways the series could last six games. When he could not insert any more 1s between the 00, Robert moved the placement of 00 from first two

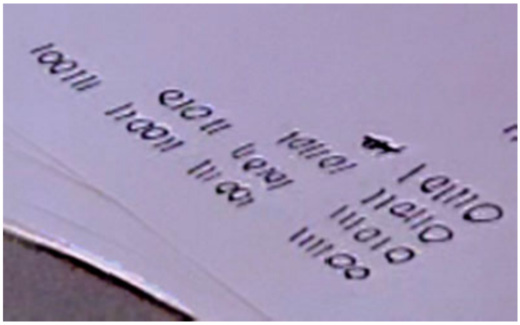


Fig. 10 Robert's binary list for all possible six-game series

spots to second and third spot (second column in his list) and repeated the pattern of inserting 1s in between the 00. Once Robert had placed 00 in fourth and fifth spot, he knew he was done. He indicated that team 1 had to win the sixth game for the series to run six games. He carefully did not write the cases where the game could end sooner than six games (like 111100). Once his list was complete, Robert was confident that there were 20 ways for the Series to end in six games, as there were 10 ways for each of the teams to win. After Michael and Robert agreed that there were 20 ways the series could in six games, they started to work on the seven-game series.

Michael suggested to Robert that the combinations formula might be helpful and Robert agreed. Michael worked on the six-game series and calculated that $C(6,4) = 15$ so there are 30 ways for the game to end as each team could win in 15 possible ways. However, Michael explained that this 30 also includes the 10 ways that the series could have ended in five games, given by $C(5,4)$. To correct this over counting, they needed to subtract 10 from 30 giving the 20 possibilities that Robert's binary listed had previously indicated. Next, Michael pointed out that they should perhaps subtract a 2 from this 20 to take care of the cases where the series ends in four games. Robert quickly dismissed this suggestion and explained that the 10 outcomes for the five-game case already included the 2 cases for the four-game series and therefore, did not need to be subtracted. Robert, relying on his control for a variable technique, was convinced that 20 were the correct number of possibilities. Michael agreed with Robert after some pondering. The list that Robert created for the six-game series deserves an interesting digression. We share here an earlier episode where Robert had used this technique for controlling a variable prior to the discussion of the World Series problem as it provides another example of how students recycled and revised their representations to find solutions for the new counting problem tasks.

5.6 Controlling for a Variable

In the beginning of his 11th grade, Robert worked on a problem where he was asked to find all five-tall towers that have exactly two blue blocks in them when choosing

Fig. 11 Robert's arrangement of six-tall towers with exactly two blue blocks

from blue and yellow colors. Robert worked with another student Michelle in this session.

Robert first guessed that for two-tall towers, there are two towers containing exactly two blue cubes, and that there are four towers with exactly two blues for the three-tall case, followed by six such towers in the four-tall case. Robert was trying to build an inductive argument and conjectured that there would be 8 towers that are five-tall with exactly two blues in them. However, he soon realized that for the three-tall towers, there are two towers that meet the requirement and not four as he had previously estimated. With this realization, he changed his guess to 10 towers for the five-tall case. To convince themselves further, he and Michelle built the towers five tall with two blues and settled with 10 towers as the solution. Then the researcher, Maher, asked students to think of the six-tall towers with exactly two blues in them.

To create all the six-tall towers, Robert used the strategy of controlling for a variable, this time the blue color. Robert started with placing two blue (dark) blocks on the top of the tower and starting inserting yellow (light) blocks in them. He increasingly inserted yellow blocks till he could not insert any more creating towers as follows: BBYYYY, BYBYYY, BYYBYY, BYYYBY, and BYYYYB where B is blue and Y is yellow (Fig. 11). Next, he moved the two blue blocks down by one position and repeated the insertion of yellow blocks. This process lead Robert to find 15 towers that are six-tall with exactly two blues in them. As detailed earlier, Robert easily modified this technique to list the ways a six-game series could end and was confident of the sample space he listed for both the towers problem and the World Series problem. This instance reflects the strength of meaningful ideas that students built earlier and owned, when given the opportunity and time to work on and revisit the structured problem solving tasks.

In this episode, Robert decided to investigate patterns in the towers of a certain height that meet the criteria of exactly two blues in them. He recorded his results in a table as shown below. Robert first discovered a vertical pattern in the table and realized that the entry in the following row was found by adding the next positive integer, that is, 0 from first row $+1 = 1$ in the second row, 1 from the second row $+2 = 3$ in the next row, and so on.

Height of the tower	Number of towers with exactly two blues
1	0
2	1
3	3
4	6
5	10
6	15

However, Robert stated that he needed to find the pattern across since he would be unable build on the vertical pattern if the height of the towers grew considerably. After some more time pondering over the pattern in this table, Robert later discovered that for a height h of the tower, the number of towers is give by $h(h/2 - 0.5)$ which simplifies to $h(h - 1)/2$ or the formula for $\mathrm{C}(h, 2)$. Robert had unintentionally but quite impressively discovered the formula for $\mathrm{C}(h, 2)$ that he later used to solve other counting tasks, like the problem of choosing a committee of two out of five people, and made several interesting connections to the Pascal's Triangle (Ahluwalia 2011).

6 Conclusions and Implications

Through the episodes shared in this chapter, we were able to observe how students, young and in later years, worked collaboratively to build mathematical ideas and create novel representations when given an opportunity to work on challenging tasks with minimum intervention. The students naturally learned mathematical ideas from each other and often created and adopted universal representations when appropriate and convincing arguments were presented (Maher 2010). For example, to verify fairness of the dice game, students accepted Stephanie's histogram as a correct representation for the outcomes for rolling a pair of dice after they were convinced that the events $(2, 1)$ and $(1, 2)$ should be counted as two different events that lead to the sum of three. As another example, Milin's inductive argument was accepted and assimilated by several students and used by them over the years to explain the tree-diagram method for building towers (Sran 2010). Also, the binary list that was used by Michael and Robert in the episodes documented here was initially introduced by Michael in solving the pizza problem in grade 6 (Muter and Uptegrove 2010).

We also see that in attempts to organize and list the outcomes of a probabilistic experiment, students (Stephanie, Michelle and Ankur) naturally create tree diagrams; use sophisticated lists (Robert and Michael); and sometimes invent mathematical formulae (Ankur, Robert). Given the opportunity to work collaboratively on meaningfully designed tasks with ample time to revisit and review their ideas, students are able to build on their mathematical justifications and representations that become more sophisticated and elaborate over time (Maher et al. 1993). Also, we see that these representations are flexible enough to grow with and accommodate the structure of increasingly challenging tasks. For example, Robert used the tree diagram as a sophisticated mathematical organizer that enabled him to connect his older counting ideas to the mathematically advanced ideas related to the structure of Pascal's Pyramid (Ahluwalia 2011).

We contend that the ideas that students build while working on meaningful counting tasks provide a crucial and effective foundation for later probabilistic reasoning. In their investigations of counting tasks, the students learned to correctly and convincingly find sample spaces, an important first step in determining the probability of an event. As Kiczek (2000) demonstrated in her analysis of the students' World Series problem-solving, Ankur, Jeff, Romina, Michael, and Brian were able to correctly find the solution by folding back to their basic understanding of other counting tasks like the pizza problem and the building towers activity. The students not only found the sample space for the different cases of the series, but also were able to calculate the probabilities for each event with reliable justifications. Furthermore, the students built connections with Pascal's Triangle, towers, the binary digits, and pizzas problem to convince themselves that they had not overlooked any possibilities. Recognizing the connections between and among the various representations was an effective strategy for the students to monitor the validity of their problem-solving reasoning.

In some cases, students were able to use their basic blocks of understanding along with their hands-on experience with the counting tasks to discover new combinatorial formulae. For example, Robert discovered the combinations formula while working on a specific towers activity (Ahluwalia 2011). Robert was later able to reuse and justify this formula in several other counting tasks and adapted it expertly to find sample size of challenging counting tasks (Ahluwalia 2011).

The students in the longitudinal study had opportunities over the years to work on challenging counting tasks. They demonstrated their understanding by creating powerful mathematical arguments and justifications to support their work. Later, when the students were invited to solve probabilistic tasks, they folded back on their robust counting schemes to find the sample spaces, test the fairness of the events, and correctly calculate the probabilities. It is our view that when students are invited to think and act like mathematicians, that is, build a strong foundational understanding to guide their intuition, they develop ownership and create meaningful representations for their ideas. This knowledge serves them well in working collaboratively with their peers, to negotiate solutions and justifications, providing a strong foundation for a rewarding lifelong learning of mathematics. We highly recommend that the opportunities and environments that were experienced by the longitudinal study students be routinely provided to all students across all age groups. Every student should engage in problem-solving investigations in the counting domain in their early school years as a part of their mathematical journey to build a rich foundation for exploring more sophisticated mathematical concepts that they will encounter in later years.

References

Ahluwalia (2011). *Tracing the building of Robert's connections in mathematical problem solving: a sixteen-year study*. Unpublished PhD dissertation, Rutgers University, NJ.

Alston, A., & Maher, C. A. (2003). Modeling outcomes from probability tasks: sixth graders reasoning together. In N. Pateman, B. Dougherty, & J. Zilliox (Eds.), *Proceedings of the joint meeting of the 27th annual international group and 25th North American group of the psychology of mathematics and education*, Honolulu, HI, USA (Vol. 2, pp. 2–32).

Amit, M. (1998). Learning probability concepts through games. In L. Pereira-Mendoza, L. S. Kea, T. W. Kee, & W. K. Wong (Eds.), *Proceedings of the international conference on the teaching of statistics (ICOTS-5)*, Singapore (pp. 45–48).

Benko, P. (2006). *Study of the development of students' ideas in probability*. Unpublished doctoral dissertation, Rutgers University, NJ.

Benko, P., & Maher, C. A. (2006). *Students constructing representations for outcomes of experiments*. Paper presented at the thirtieth annual meeting of the international group for the psychology of mathematics education, Prague, Czech Republic (Vol. 2, pp. 137–144).

Dann, E., Pantozzi, R. S., & Steencken, E. (1995). Unconsciously learning something: a focus on teacher questioning. In *Proceedings of the 17th annual meeting of the psychology of mathematics education*. Columbus: North American Chapter.

Davis, R. B., & Maher, C. A. (1990). What do we do when we do mathematics? In R. B. Davis, C. A. Maher, & N. Noddings (Eds.), *Constructivist views on the teaching and learning of mathematics*. *Journal for research in mathematics education monograph*, *4*, 65–78. Reston: National Council of Teachers of Mathematics.

Davis, R. B., & Maher, C. A. (1997). How students think: the role of representations. In L. D. English (Ed.), *Mathematical reasoning: analogies, metaphors, and images* (pp. 93–115). Hillsdale: Erlbaum.

Fainguelernt, E., & Frant, J. (1998). The emergence of statistical reasoning in Brazilian school children. In L. Pereira-Mendoza, L. S. Kea, T. W. Kee, & W. K. Wong (Eds.), *Proceedings of the international conference on the teaching of statistics (ICOTS-5)*, Singapore (pp. 49–52).

Francisco, J. M., & Maher, C. A. (2005). Conditions for promoting reasoning in problem solving: insights from a longitudinal study. *The Journal of Mathematical Behavior*, *24*, 361–372.

Kiczek, R. D. (2000). *Tracing the development of probabilistic thinking: profiles from a longitudinal study*. Unpublished doctoral dissertation, Rutgers University.

Kiczek, R. D., & Maher, C. A. (2001). *The stability of probabilistic reasoning*. Paper presented at the twenty-third annual conference of the North American Chapter of the international group for the psychology of mathematics education, Snowbird, Utah.

Kiczek, R. D., Maher, C. A., & Speiser, R. (2001). Tracing the origins and extensions of Michael's representation. In A. A. Cuoco & F. R. Curcio (Eds.), *The role of representation in school mathematics NCTM yearbook 2001* (pp. 201–214). Reston: National Council of Teachers of Mathematics.

Maher, C. A. (1998). Is this game fair? The emergence of statistical reasoning in young children. In L. Pereira-Mendoza, L. S. Kea, T. W. Kee, & W. K. Wong (Eds.), *Proceedings of the international conference on the teaching of statistics (ICOTS-5)*, Singapore.

Maher, C. A. (2002). How students structure their own investigations and educate us: what we have learned from a fourteen-year study. In A. D. Cockburn & E. Nardi (Eds.), *Proc. 26th conf. of the int. group for the psychology of mathematics education* (Vol. 1, pp. 31–46). Norwich: University of East Anglia.

Maher, C. A. (2005). How students structure their investigations and learn mathematics: insights from a long-term study. *The Journal of Mathematical Behavior*, *24*(1), 1–14.

Maher, C. A. (2008). The development of mathematical reasoning: a 16-year study. In M. Niss (Ed.), *Proc. of 10th int. congress for mathematics education*. Roskilde: Roskilde University.

Maher, C. A. (2010). The longitudinal study. In C. A. Maher, A. B. Powell, & E. B. Uptegrove (Eds.), *Combinatorics and reasoning: representing, justifying, and building isomorphisms* (pp. 3–8). New York: Springer.

Maher, C. A., & Martino, A. M. (1996a). The development of the idea of mathematical proof: a 5-year case study. *Journal for Research in Mathematics Education*, *27*(2), 194–214.

Maher, C. A., & Martino, A. M. (1996b). Young children invent methods of proof: the gang of four. In P. Nesher, L. P. Steffe, P. Cobb, B. Greer, & J. Golden (Eds.), *Theories of mathematical learning* (pp. 431–447). Mahwah: Erlbaum.

Maher, C. A., & Martino, A. M. (1997). Conditions for conceptual change: from pattern recognition to theory posing. In H. Mansfield & N. H. Pateman (Eds.), *Young children and mathematics: concepts and their representations*. Sydney: Australian Association of Mathematics Teachers.

Maher, C. A., & Martino, A. M. (2000). From patterns to theories: conditions for conceptual change. *The Journal of Mathematical Behavior*, *19*(2), 247–271.

Maher, C. A., & Muter, E. M. (2010). Responding to Ankur's challenge: co-construction of argument leading to proof. In C. A. Maher, A. B. Powell, & E. B. Uptegrove (Eds.), *Combinatorics and reasoning: representing, justifying and building isomorphisms* (pp. 9–16). New York: Springer. Chapter 8.

Maher, C. A., & Speiser, R. (2002). The complexity of learning to reason probabilistically. In *Proceedings of the twenty fourth annual meeting of the North American Chapter of the international group for the psychology of mathematics education*, Columbus, Ohio (Vols. 1–4).

Maher, C. A., & Uptegrove, E. B. (2010). Methodology. In C. A. Maher, A. B. Powell, & E. B. Uptegrove (Eds.), *Combinatorics and reasoning: representing, justifying and building isomorphisms* (pp. 9–16). New York: Springer. Chapter 2.

Maher, C. A., & Weber, K. (2010). Representation systems and constructing conceptual understanding. *Mediterranean Journal for Research in Mathematics Education*, *9*(1), 91–106.

Muter, E. M., & Uptegrove, E. B. (2010). Representations and connections. In C. A. Maher, A. B. Powell, & E. B. Uptegrove (Eds.), *Combinatorics and reasoning: representing, justifying and building isomorphisms* (pp. 105–120). New York: Springer. Chapter 10.

Maher, C. A., Martino, A. M., & Alston, A. S. (1993). Children's construction of mathematical ideas. In B. Atweh, C. Kanes, M. Carss, & G. Booker (Eds.), *Contexts in mathematics education: Proc. of 16th annual conf. of the mathematics education research group of Australia (MERGA)*, Brisbane, Australia (pp. 13–39).

Maher, C. A., Powell, A. B., & Uptegrove, E. B. (Eds.) (2010). *Combinatorics and reasoning: representing, justifying and building isomorphisms*. New York: Springer.

National Council of Teachers of Mathematics (2000). *Principles and standards for school mathematics*. Reston: NCTM.

Pirie, S., Martin, L., & Kieren, T. (1996). Folding back to collect: knowing what you need to know. In *Proceedings of the 20th annual conference of the international group for the psychology of mathematics education*, Valencia, Spain.

Powell, A. B., Francisco, J. M., & Maher, C. A. (2003). An analytical model for studying the development of learners' mathematical ideas and reasoning using videotape data. *The Journal of Mathematical Behavior*, *22*(4), 405–435.

Shay, K. B. (2008). *Tracing middle school students' understanding of probability: a longitudinal study*. Unpublished PhD dissertation, Rutgers University, NJ.

Speiser, R., & Walter, C. (1998). Two dice, two sample spaces. In L. Pereira-Mendoza, L. S. Kea, T. W. Kee, & W. K. Wong (Eds.), *Proceedings of the international conference on the teaching of statistics (ICOTS-5)*, Singapore (pp. 61–66).

Sran, M. K. (2010). *Tracing Milin's development of inductive reasoning: a case study*. Unpublished PhD dissertation, Rutgers University, NJ.

Vidakovic, D., Berenson, S., & Brandsma, J. (1998). Children's intuitions of probabilistic concepts emerging from fair play. In L. Pereira-Mendoza, L. S. Kea, T. W. Kee, & W. K. Wong (Eds.), *Proceedings of the international conference on the teaching of statistics (ICOTS-5)*, Singapore (pp. 67–73).

Levels of Probabilistic Reasoning of High School Students About Binomial Problems

Ernesto Sánchez and Pedro Rubén Landín

Abstract In this chapter, some aspects of the process in which students come to know and use the binomial probability formula are described. In the context of a common high school probability and statistics course, a test of eight problems was designed to explore the performance of students in binomial situations. To investigate the influence of instruction to overcome some common cognitive bias or their persistency, the first three problems are formulated in a way that may induce bias. Each one is structurally equivalent to another problem phrased to avoid any bias that was included in the test. Also, the second and third problems were administered before and after the course to assess the changes produced by instruction. A hierarchy of reasoning, designed in a previous study, was adapted and used to classify the answers of the students in different levels of reasoning. The classification of these answers points out that the components of knowledge, the classical definition of probability, the rule of product of probabilities, combinations, and the binomial probability formula, are indicators of transitions between levels. The influence of the phrasing of the problems is strong before instruction, but weak after it.

1 Introduction

The binomial distribution is one of the most important discrete probability distributions. It is part of the high school syllabus and also of the first college course of statistics. The probability content of high school statistics courses includes probability distributions of elementary random variables (at least binomial and normal), mathematical expectation and variance. As Jones et al. comment, "The major new conceptual developments at this level are the inclusion of random variable and probability distributions" (Jones et al. 2007, p. 914).

On the other side, it has been found that people respond to certain kind of probability problems influenced by some cognitive bias, like representativeness, availability, the illusion of linearity, the supposition of equiprobability, etc. (Fischbein and Schnarch 1997; Batanero and Sánchez 2005). Students respond to several situations

E. Sánchez (✉) · P.R. Landín
Cinvestav-IPN, México, DF, Mexico
e-mail: esanchez@cinvestav.mx

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_31,

based on heuristic of representativeness when it is ineffective in the given cases, and that behavior seems to be persistent even after students have learned some techniques of probability which would help them avoid it (Shaughnessy 1992). However, despite the importance to clarify this observation by Shaughnessy, there is no research report that describes and analyzes how students who have learned some techniques of probability contend with problems that may influence them to respond based on cognitive bias. Are the answers of these students similar to those of students who have not studied probability?

However, for purposes of teaching, it is not only important to reveal the relationships between learning and cognitive bias but also to understand how probability knowledge is acquired by students throughout a course of probability. Hierarchies are a way to describe knowledge acquisition. Indeed, by using a hierarchy it is intended to describe how and in which order students learn certain content knowledge (Hart 1981). In this paper, a hierarchy on binomial formula is proposed to answer the question: How do the students learn the binomial probability formula?

2 Background

There are two kinds of studies concerned with the present work; one of them is related to the elaboration of probabilistic reasoning hierarchies, and the other to the learning of the binomial probability formula.

Reading and Reid (2010) argue that hierarchies can be considered cognitive frameworks that are useful for (i) informing the sequencing and presentation of content in the curriculum, (ii) designing learning activities to best support student learning, (iii) elaborating assessment tasks, and (iv) judging whether an assessment task is eliciting a cognitively sound representation of reasoning.

They mention two examples of cognitive frameworks for statistical reasoning which are based on the 'Structure of Observed Learning Outcome' (SOLO) Taxonomy. The first one, proposed by Langrall and Mooney (2002), describes levels achieved by students during four different processes, each of these corresponding to basic probability concepts. The *idiosyncratic, transitional, quantitative*, and *analytical* levels are proposed for this framework. The other framework, proposed by Garfield (2002), was designed to describe the levels of reasoning achieved by students working with statistical concepts. The levels are: *idiosyncratic, verbal, transitional, procedural*, and *integrated*. These studies show how cognitive levels are described to inform teaching and research.

SOLO Taxonomy has been frequently used to analyze or assess the understanding of statistics or probabilistic concepts resulting in a hierarchy. For example, Tarr and Jones (1997) proposed a hierarchy about conditional probability and independence; Jones et al. (1999) extended it to cover sample space, theoretical and experimental probability; and comparing probabilities. Watson and Moritz (1999a) studied groups comparing tasks and Watson and Moritz (1999b) analyzed the average concept. Reading and Shaughnessy (2004) and Reading and Reid (2007) offer hierarchies for variation reasoning. Aoyama (2007) proposed a hierarchy to graph interpretation. Reading and Reid (2006) studied the reasoning about and with the distribution concept. These authors proposed a hierarchy for empirical distributions; however, particular theoretical distributions such as the binomial one should also be analyzed from the same perspective. For this purpose and due to the complexity of the binomial distributions, we started to investigate the reasoning about and with the binomial probability formula, leaving the wider concept of binomial distribution for future studies.

In our review of the literature, we did not find studies in which a hierarchy for binomial formula or binomial distribution is proposed; moreover, there are a few empirical researches on the learning of these concepts. In contrast, there are many proposals for the instruction of the binomial probability formula or distribution (e.g., Rouncefield 1990; Gross 2000; Chalikias 2009). With respect to the empirical studies, those of Van Dooren et al. (2003) and Abrahamson (2009a, 2009b) are briefly described.

Van Dooren et al. (2003) explore the presence of the *illusion of linearity* phenomenon when students respond to problems that can be solved with the knowledge of some techniques and concepts involved in the binomial formula. The illusion of linearity phenomenon is the overgeneralization of properties of a lineal function; in many situations the proportionality is considered to be the tool to solve the problem. Problems like the following were administered before and after instruction to 225, 10th (aged 15–16) and 12th (aged 17–18) grade secondary school students: Say if the following statement is true or false: *I roll a fair die several times. The chance to have at least two times a six if I can roll* 12 *times is three times as large as the chance to have at least two times a six if I can roll four times*. The authors conclude that:

> The large majority of students chose for the response alternative that stated a linear increase (or decrease) of the probability of the described event if one or two variables in the situation increased (or decreased). A further, more fine-grained analysis of the erroneous answers indicated that the large majority of errors could indeed unequivocally be characterized as resulting from students' over-reliance on the linear model (p. 133).

Three case studies of undergraduate students reasoning about a simple probability situation (binomial) were examined by Abrahamson (2009a) to show how a semiotic approach illuminates the process and content of students reasoning.

He asked students to analyze an experimental procedure in which four marbles were randomly drawn out from a box containing hundreds of marbles. In the box were equal amounts of green and blue marbles. The draws were undertaken with a *marble scooper*, a device with four concavities arranged in a 2-by-2 array. The features of the marble scooper help students represent all possible results of the experiment in a 2-by-2 array, and then, with the help of a computer program the students were guided to construct the distribution of the variable "number of green marbles". Abrahamson shows how the knowledge is appropriated by students using different resources. Students objectify their ideas for the sake of reasoning and communication triggered by semiotic instruments.

Abrahamson (2009b) reports an interview with an 11.5-year old student to find out how students make sense of topics of the binomial distribution they are taught, specially the results of flipping a coin four times. Two artifacts of mediation were used: The marble scooper and cards with an empty 2×2 matrix. These cards were used to register the elements of the sample space and organize an array similar to the distribution of the variable ‘number of green marbles’. Student overcame the belief that the values of this random variable are equiprobable and calculated the corresponding probabilities with understanding.

Landín and Sánchez (2010) proposed a hierarchy to assess the answers of high school students to tasks on binomial distribution. The hierarchy could be interpreted as a hypothetical learning path of knowledge components of the binomial probability formula; the authors also suggest that the binomial structure of the problems can be hidden behind the phrasing of the problem statements when these may be misinterpreted based on some cognitive bias. Sánchez and Landín (2011) describe the process followed to give greater reliability to that hierarchy of reasoning and improve it. In the present work, the test to gather data has been improved with the purpose of clarifying some hypothesis emerged in the previous studies.

3 Conceptual Framework

The Structure of Observed Learning Outcome (SOLO) (Biggs and Collis 1982, 1991) was used in helping defining a hierarchy to account for the reasoning about binomial probability formula and analyze the answers to the test. According to Biggs and Collis (1991) five levels of abstraction or models of functioning can be distinguished to describe the development of a child: sensorimotor (from birth), iconic (from around 18 months), concrete-symbolic (from around 6 years), formal (from around 14 years), and post-formal (from around 20 years). High school students are going through the formal mode, this means that they are able to understand abstract systems in which all elements are interconnected and that can be used to generate hypotheses about new situations: “thinking in the formal mode thus both incorporates and transcends particular circumstances” (Biggs and Collis 1991, p. 63). Within each mode answers become increasingly complex and this growth can be described in terms of the SOLO-levels which are: unistructural (U) when the answer includes only one relevant component of the concept or task, multistructural (M) when it includes several unrelated components or with spurious relationships, and relational (R) when it includes several components with genuine relationships among them. Two additional levels named *prestructural and extended abstract* connect with the previous and the following mode, respectively. In the first, no component of the concept is considered; in the second, a complete domain of the components and its relationships is achieved.

We would like to emphasize two aspects of SOLO: (i) The SOLO levels are task-related since in one task a student may respond in a level and in other tasks, in different levels; so, it is not possible to assign a child to a level; (ii) “It must be

Table 1 Components of knowledge on binomial situations and the probability formula

1.	Recognition of Bernoulli situations (p, q)
2.	Recognition of the binomial random variable (k)
3.	Use of combinatorial trees and representation of sequences of Es and Fs (size n)
4.	Use of classical probability definition (Laplace definition)
5.	Knowledge and use of product rule of probabilities
6.	Use of combinations and product rule of probabilities without integrating them in the formula
7.	Use of binomial probability formula

stressed that it is what is *observed* in students' responses that is analyzed, not what the observer thinks the student might have meant" (Watson 2006, p. 13).

The procedure that we follow to define a hierarchy for the binomial probability formula was to make *a priori* analysis of the conceptual components underpinning the knowledge of binomial distribution (Table 1), and with those components define SOLO levels. After that, the hierarchy was elaborated taking into account answers of the students to a preliminary test and a process to improve the reliability of the hierarchy.

A problem emerged when answers were classified according to the early hierarchy we defined in Landín and Sánchez (2010). The answers of most students were rather messy and sometimes contain errors, for example, an answer could show that a student used two or more components of knowledge to try to solve the problem. Then, the answer would be in multistructural level; however, the answer may also show arithmetic errors or strange elements that make it difficult to classify the answer in the same level as the answers without these elements. Considering errors important, we add performance components to define the hierarchy. The SOLO taxonomy does not prescribe what to do with errors so the hierarchy proposed is not aligned with it; that is the reason why we prefer to number levels from 1 to 5 instead of using terminology of SOLO levels.

4 Methodology

In this paper, we present the results of analyzing the answers to a test of 8 problems and two problems of a pretest, most of them about the binomial formula, with a hierarchy of reasoning. Two studies were previously undertaken (Landín and Sánchez 2010; Sánchez and Landín 2011) where the first steps to construct the hierarchy of the binomial formula were given. In the first study, an *a priori* analysis of the conceptual elements underpinning the knowledge of binomial distribution (components of knowledge) and the SOLO levels were used to design a hypothetical hierarchy of reasoning. Then, a test with four binomial problems was made and applied to 66 high school students (17–18 years old) after they participated in a probability course with emphasis on binomial concepts. Finally, the answers were analyzed and categorized according to the hierarchy.

In the second study, we described the process followed to increase the reliability of the hierarchy: Six specialists were asked to apply the hierarchy on the corpus of answers obtained in the previous study. The comparison and analysis of the classifications made, especially in the answers where there was no consensus, led to restructuring the hierarchy. Performance components were added to knowledge components as criteria to make a decision about where an answer should be classified. In the current study, the test used in the previous studies was restructured, administered to a new group of students, and the results were classified with the restructured hierarchy.

Participants Twenty six students participated in a probability course as part of their last year of high school. They had not previously studied any course on this subject. They were administered tests during the course and were not informed about the investigative intention of them; in fact, the tests were part of the activities to grade students' performance during the course.

Instruments Two tests with two common items were elaborated with the intention of gathering information concerning the students reasoning about components of binomial probability formula. The first has four problems. The second one has eight problems: three are phrased in such way that they may induce students to answer based on some cognitive bias; another three are structurally equivalent to the first three but the questions are formulated in a direct way to avoid any confusion. The remaining two are almost the same problem, except that one asks for the calculation of one value of the variable while the other asks for several using the expression "at least 3 shots". With respect to the first test, only two questions are analyzed and they are identical to two problems of the second test (Problems 2 and 3).

Procedure Data of this study were obtained from an ordinary course of probability and statistics in a public high school in Mexico. The teacher was one of the authors (PRL) and he had to follow the syllabus introducing all topics of probability in five 50-minute weekly sessions during fourth months: sample space, events and definitions of probability; combinatorics, conditional probability and independence; and discrete random variables and their distributions. Instruction was developed combining lectures and problem solving activities. Frequentists' approach of probability was reinforced with simulation activities using *Fathom* software, and some binomial problems were solved with simulation techniques and their solutions were compared to those found with calculations.

During the development of the course, the two tests analyzed here were administered. The first test is considered a pre-test since it was administered before the introduction of binomial problems in class. The second one was administered after the students had studied the topic on the binomial distribution; it permitted assessing their learning.

The Hierarchy The hierarchy we proposed in Sánchez and Landín (2011) is described in Table 2.

Table 2 Levels of reasoning for the binomial probability formula

Level 1.	Answers are idiosyncratic or influenced by cognitive bias, as representativeness, the illusion of linearity, the incorrect supposition of equiprobability, or other reasons. They may present traits of knowledge components but without coherence or with a great deal of errors.
Level 2.	Answers are based on a description of the sample space, possibly incomplete, and the use of the classical probability definition. The product rule might be used but with errors or partial operations. Combinations are calculated by making lists or using the formula but with some errors.
Level 3.	Correct answers are characterized by the use of combinatorial procedures and the classical definition of probability, generally supported by tree diagrams. The product rule might be used to obtain partially correct answers. Pascal's triangle can be used to calculate combinations; in some cases binomial coefficients are used but with incorrect calculations. In some answers, the binomial probability formula may be written but without really being used.
Level 4.	Correct answers are characterized by the use of product rule and combinatorial procedures, mostly accompanied by tree diagrams. The combinations formula is used to calculate binomial coefficients. Answers that are partially right or that have some errors might be obtained with the binomial probability formula.
Level 5.	The correct use of the binomial formula is showed at this level. The values of the parameters are identified and substituted in the formula and the calculations are well done.

5 Results

Answers to Problems 2, 3, and 4 of the test were chosen to exemplify the kind of answers that lie in the different levels of the hierarchy. The three problems require a binomial model to reach their solution. Problems 2 and 3 are also problems of the pre-test and are useful for comparing the performance of the students before and after the instruction; they come in pairs in the post-test: 2–6 and 3–7, where the phrasing of the first problem of each pair may induce to apply a cognitive bias. This is the reason why they were chosen.

Level 1 It is similar to the *subjective* level of Jones et al. (1999) and the *prestructural* level of Biggs and Collis (1991). In the example of Problem 2, given in Fig. 1, a student's answer is guided by representativeness (Student 9).

Problem 2 (Product rule) The probability of having a newborn girl is the same as having a newborn boy and they are equal to $1/2$. In a family of five children, which of the following statements is true?

(a) The event BGBGG is more probable than the event BBBBB.
(b) The event BBBBB is more probable than the event BGBGG.
(c) Both events have the same probability.
BGBGG means that the first child is a boy, the second a girl, the third a boy, and so on.

Justify your choice.

The probability of both (B and G) is 50 % and given that the result has a certain variability, it is more probable that in a family there are (were) women and men than only men or only women.

Fig. 1 Answer to Problem 2 in level 1

Fig. 2 Answer to Problem 3 in level 1

The events have the same probability because they are equivalent:

1 head when 2 coins are flipped $= (1/2) = 0.5$;

2 heads when 4 coins are flipped $= (2/4) = (1/2) = 0.5$.

Fig. 3 Answer to Problem 4 in level 1

Problem 4 (14)	$\sigma_x = \sqrt{pq}$,
$P_e = 0.7$ %,	$\sigma_x = \sqrt{(0.7)\,(0.3)}$,
$P_f = 0.3$ %,	$\sigma_x = \sqrt{0.21} = 0.4582$.
$P_e = 1$,	
$\mu = p$,	
$\mu = 0.7$,	

The procedure showed in the answer to question 3 (Fig. 2) is based on the illusion of linearity (Student 13).

Problem 3 (Comparisons of binomial probabilities) What is more probable?

(a) Getting one head when two coins are flipped;
(b) Getting two heads when four coins are flipped;
(c) They are equally probable.

Justify your choice.

Figure 3 in the answer to question 4 indicates a calculation similar to that of the standard deviation of the binomial distribution, but it is not pertinent to the problem (Student 9).

Problem 4 (Common problem of binomial probability) A shooter has a probability of 0.7 to hit the target and 0.3 to fail. What is the probability of hitting the target exactly 3 times out of 5 shots?

Level 2 It is similar to the *transitional* level of the hierarchy of Jones et al. (1999) and in some aspects to the *unistructural* of Biggs and Collis (1991). The answer of Student 18 to question 2 (Fig. 4) shows that a tree diagram was used to calculate the elements of the sample space, however, the student calculated the probability of the event "two boys and 3 girls in a family" instead of the ordered result "BGBGG".

Two boys and 3 girls are more probable; since its probability is $(7/32) = 0.21$. There is only a case for "all are boys", therefore there are more cases for 2 boys and 3 girls.

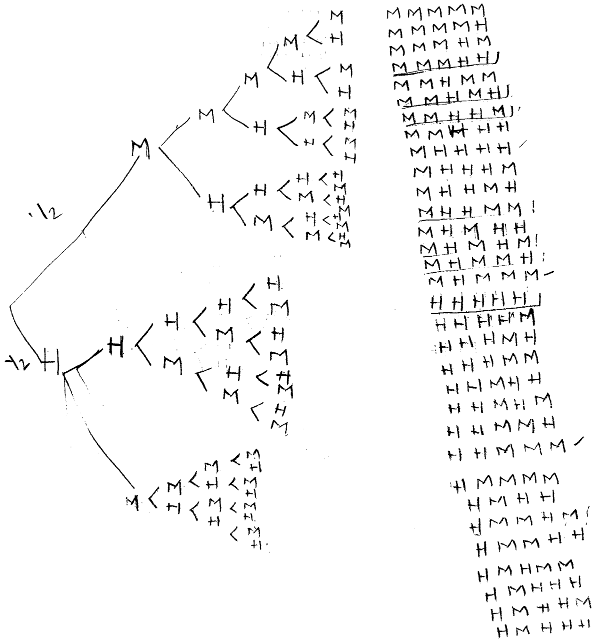


Fig. 4 Answer to Problem 2 in level 2

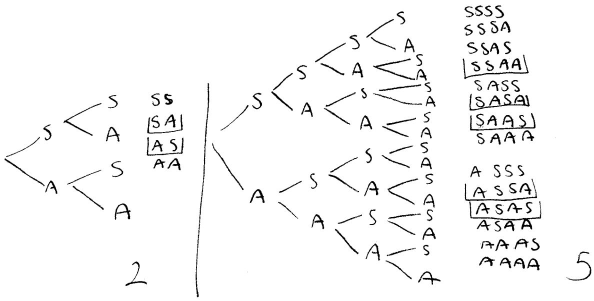


Fig. 5 Answer to Problem 3 in level 2

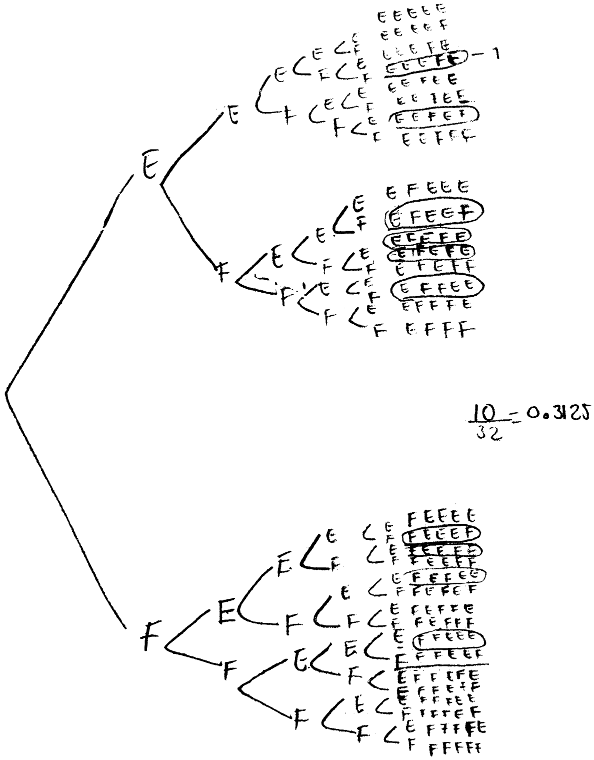
In Fig. 5, Student 7 constructed a tree diagram for the two events, and possibly compared the favorable cases of each experiment, concluding incorrectly that the right answer was option (b).

The answer of Student 12 (Fig. 6) consisted of a tree diagram used to count combinations and used to calculate the probability of the events; however, she supposed that E and F had the same probability, but they did not.

Level 3 This is similar to the *informal quantitative* level of Jones et al. (1999) and *the multistructural* level of Biggs and Collis (1991).

Student 5 (Fig. 7) constructed a tree diagram of probability, identified the outcomes to be compared in it, and calculated their probabilities using the product rule or the Laplace definition.

Fig. 6 Answer to Problem 4 classified in level 2



Student 6 (Fig. 8) constructed two probability tree diagrams, identified the favorable outcomes in them, and calculated their probabilities by the Laplace definition or he might have combined the product and addition rules of probabilities. [This student is not classified in the numerical scale because in the procedure he did not explicitly apply the product rule.]

Fig. 7 Answer to Problem 2, classified in level 3

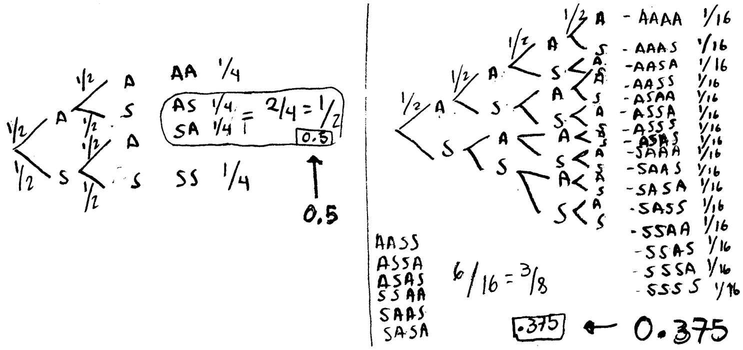


Fig. 8 Answer to Problem 3, classified in level 3

EFEFE = 0.03087
E = 0.7
F = 0.3
"Supposing that 'EFEFE' is used, since it is an experiment where there are 5 shots, the following operation is made: E × F × E × F × E, where E = 0.7 and F = 0.3; substituting: (0.7)(0.3)(0.7)(0.3)(0.7) = 0.03087".

Fig. 9 Answer to Problem 4, classified in level 3

Fig. 10 Answer to Problem 2, classified in level 4

HHHHH = $(1/2)^5$ = (1/32) = 0.03125;
HMHMM = $(1/2)^5$ = (1/32) = 0.03125.

In the answer to Problem 4 (Fig. 9), Student 25 applied the product rule. He only considered one sequence of 3 Es and 2 Fs, but he did not take all combinations of 3 Es and 2 Fs.

Level 4 It is similar to the *numerical* level of Jones et al. (1999) and to the relational level of Biggs and Collis (1991).

Student 6 (Fig. 10) directly used the product rule of probabilities without a tree diagram.

Student 22 (Fig. 11) constructed two probability tree diagrams, identified the favorable outcomes, and calculated the probabilities of events to be compared. It should be noted that he used functional nomenclature to indicate the probabilities of outcomes.

The answer of Student 14 to Problem 4 (Fig. 12) shows the use of a probability tree diagram and the sum and product rules.

Level 5 It is similar to the *abstract extended* level of Biggs and Collis (1991). Only some answers to Problem 4 were classified in this level (an example is shown in Fig. 13). These answers were classified in the extended abstract level because the values of the parameters n, p, k were identified, the binomial probability formula enunciated, and the calculations made correctly.

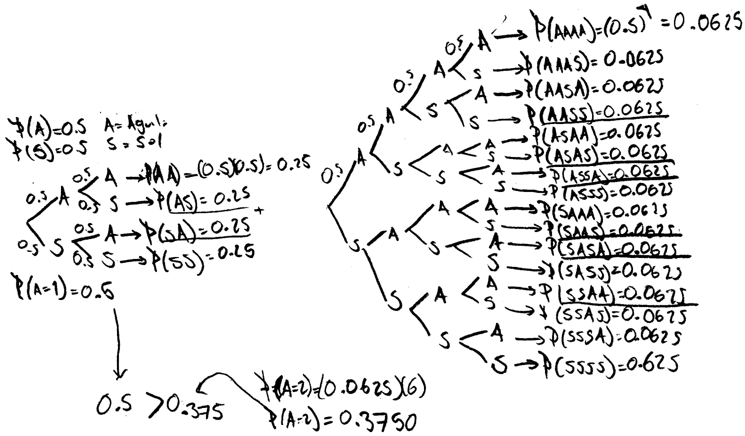


Fig. 11 Answer to Problem 3, classified in level 4

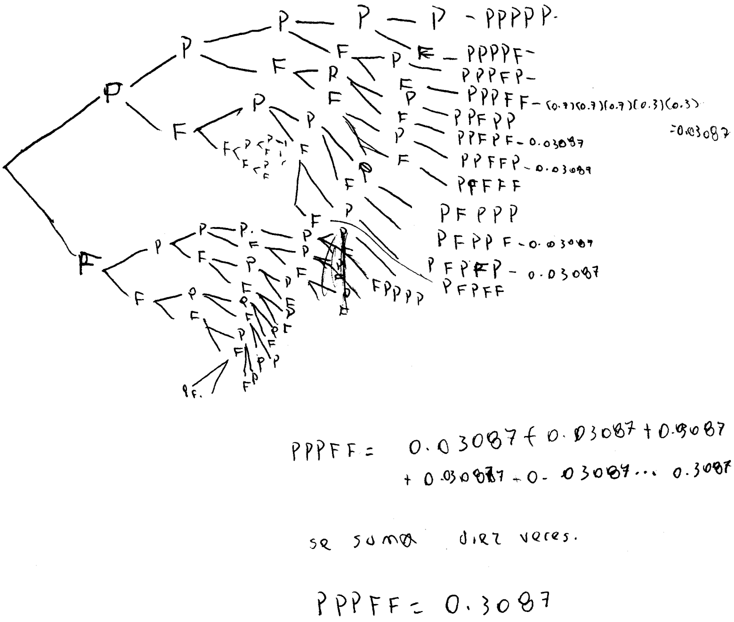


Fig. 12 Answer to Problem 4, classified in level 4

In Table 3, the frequencies of answers of the first 4 questions are presented, sorted by level. The first three problems might induce biased answers. Frequencies show that in those problems more than 1/3 of the answers lie in the subjective level of reasoning. Students found more difficulties in solving the combinatorics problem since a few answers were classified above the transitional level, while more than 1/3 in the Problems 2 and 3 lie above this level. Ten answers to the fourth problem

$n = 5,$
$k = 3,$
$p = 0.7,$
$q = 0.3,$
$P\,(x = k) = C_k^n p^x q^{n-x},$
$P\,(x = 3) = C_3^5 0.7^3 0.3^{5-3},$
$P\,(x = 3) = 10(0.34)(0.09),$
$P\,(x = 3) = 0.3.$
"He has a probability of 0.3 of hitting the target exactly 3 times out of 5 shots."

Fig. 13 Answer to Problem 4 classified in level 5

Table 3 Frequencies of answers by level of the 4 first problems of the test

Problem	Frequency of answers by level of reasoning					No answer
	Level 1	Level 2	Level 3	Level 4	Level 5	
Problem 1	11	8	2	1	2	2
Problem 2	9	4	8	5	0	0
Problem 3	9	3	11	3	0	0
Problem 4	3	6	7	4	6	0

were classified in level 4 or 5. This could be an effect of having a question free of bias.

These are the other five problems of the final test:

Problem 1 (Combinations) Committees are going to be formed from a set of 6 candidates. Are there more, less, or equal number of committees of 2 members than committees of 4 members? Mark the correct answer.

(a) More committees of 2 people than committees of 4 people can be formed.
(b) More committees of 4 people than committees of 2 people can be formed.
(c) The same number of 2 people committees and 4 people committees can be formed.

Problem 5 Committees are going to be formed out of a set of 6 candidates. Find the number of different committees of 2 people.

Problem 6 The probability of having a newborn girl is the same as the probability of having boy and they are equal to $1/2$. In a family of five children,

(a) Find the probability that all children are boys: BBBBB;
(b) Find the probability that the children of the family are three girls and two boys in the following order: BGBGG.
BGBGG means that the first child is a boy, the second a girl, the third a boy, and so on.

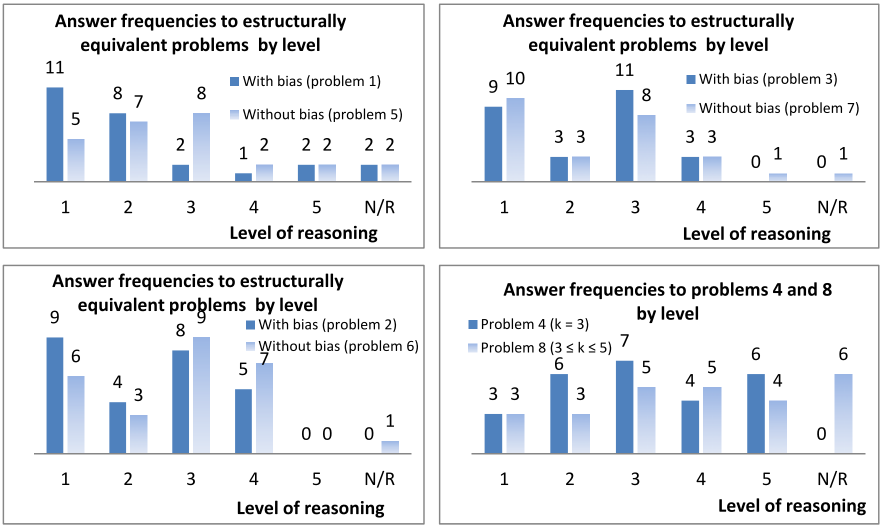


Fig. 14 Frequencies of answers to structurally equivalent problems

Problem 7 Answer each question:

(a) Two coins are flipped, what is the probability of getting exactly one head?
(b) Four coins are flipped, what is the probability of getting exactly two heads?

Problem 8 A shooter has a probability of 0.7 of hitting the target and 0.3 of failing. What is the probability of hitting the target at least 3 times out of 5 shots?

Comparing frequencies of levels of the first four problems with their structural equivalents, Fig. 14 clearly shows the influence of the statement phrasing on the solutions. The statement phrasing of questions 1 and 2 seem to influence the level of answers since those of corresponding Problems 5 and 6 were classified in higher levels of reasoning. In contrast, the answers to Problems 3 and 7 apparently had the same influence by the illusion of linearity bias. Problem 3, in which a student calculates the probability of getting 2 heads in 4 flips of a coin, is difficult by itself. It is curious that Problems 4 and 8 got better levels of answers than others; this can be explained by the phrasing without bias of Problems 4 and 8.

As expected, the effect of the course is notorious when comparing the answers to Problems 2 and 3 applied before and after the study of the binomial topic in the course (Fig. 15). The main feature that distinguishes most of the answers given before from those given after the instruction is that in the latter tree diagrams were used. This is the modal instrument to attack the problems of the test.

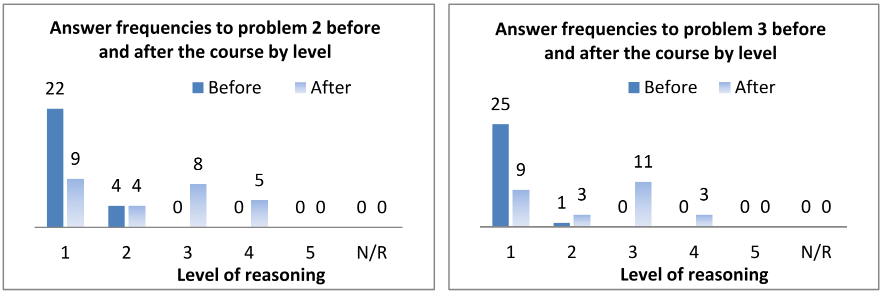


Fig. 15 Frequencies of answers before and after the course

6 Conclusions

The performance of students on binomial problems is based on the articulation of several components of knowledge, especially the product rule of probabilities and combinations; the latter in the form of arrays of Es and Fs of size n. The tree diagram seems to be a tool that makes the acquisition of those concepts viable. The comprehension of the functioning of a tree diagram helps understand the product rule of probabilities, and permits counting and having control of the number of arrays of a determined number of Es and Fs. The following step to improve the understanding of the binomial probability formula is to conceive and use the product rule and combinations. The level of reasoning which contains this understanding is achieved by a few students. The order of difficulty of the problems is: combinations, comparison of binomial probabilities, product rule, and common problems of binomial probabilities. On the other side, the phrasing of the problems influences the quality of the answers when students have not had instruction; but after it, they had difficulties due to an immature or disconnected use of components of knowledge.

Acknowledgements This work was sponsored by grant 101708 from the National Council of Science and Technology (CONACYT), México.

References

Abrahamson, D. (2009a). Embodied design: constructing means for constructing meaning. *Educational Studies in Mathematics*, *70*(1), 27–47.

Abrahamson, D. (2009b). Orchestrating semiotic leaps from tacit to cultural quantitative reasoning—the case of anticipating experimental outcomes of a quasi-binomial random generator. *Cognition and Instruction*, *27*(3), 175–224.

Aoyama, K. (2007). Investigating a hierarchy of students' interpretations of graphs. *International Electronic Journal of Mathematics Education*, *4*(3), 298–318. Retrieved from http://www.iejme.com.

Batanero, C., & Sánchez, E. (2005). What is the nature of high school student's conceptions and misconceptions about probability? In G. A. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 241–266). New York: Springer.

Biggs, J., & Collis, K. (1982). *Evaluating the quality of learning. The SOLO taxonomy* (pp. 17–31). New York: Academic Press Chap. 2.

Biggs, J. B., & Collis, K. F. (1991). Multimodal learning and the quality of intelligence behavior. In H. A. Rowe (Ed.), *Intelligence: reconceptualization and measurement* (pp. 57–76). Hillsdale: Erlbaum.

Chalikias, M. (2009). The binomial distribution in shooting. *Teaching Statistics*, *31*(3), 87–89. Retrieved from http://www3.interscience.wiley.com/journal.

Fischbein, E., & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for Research in Mathematics Education*, *28*(1), 96–105.

Garfield, J. B. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education*, *10*(3). Retrieved from www.amstat.org/publications/jse/v10n3/garfield.html.

Gross, J. (2000). Sharing teaching ideas: a Bernoulli investigation. *The Mathematics Teacher*, *93*(9), 756.

Hart, K. (1981). Hierarchies in mathematics education. *Educational Studies in Mathematics*, *12*(2), 205–218.

Jones, G. A., Langrall, C. W., & Mooney, E. S. (2007). Research in probability. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 909–955). Charlotte: Information Age-NCTM.

Jones, G. A., Thornton, C. A., Langrall, C. W., & Tarr, J. E. (1999). Understanding students' probabilistic reasoning. In L. V. Stiff & F. R. Curcio (Eds.), *Developing mathematical reasoning in grades K-12, 1999 yearbook* (pp. 146–155). Reston: National Council of Teachers of Mathematics.

Landín, P., & Sánchez, E. (2010). Niveles de razonamiento probabilístico de estudiantes de bachillerato frente a tareas de distribución binomial. *Educação Matemática Pesquisa*, *12*(3), 598–618. Retrieved from http://www.revistas.pucsp.br/index.php/emp/article/viewArticle/4842.

Langrall, C. W., & Mooney, E. S. (2002). The development of a framework characterizing middle school students' statistical thinking. In *Proceedings of the sixth international conference on teaching statistics*.

Reading, C., & Reid, J. (2006). An emerging hierarchy of reasoning about distribution: from a variation perspective. *Statistics Education Research Journal*, *5*(2), 46–68. Retrieved from http://www.stat.auckland.ac.nz/serj.

Reading, C., & Reid, J. (2007). Reasoning about variation: student voice. *International Electronic Journal of Mathematics Education*, *2*(3), 110–127.

Reading, C., & Reid, J. (2010). Reasoning about variation: rethinking theoretical frameworks to inform practice. In C. Reading (Ed.), *Data and context in statistics education: towards an evidence-based society. Proceedings of the eighth international conference on teaching statistics*, ICOTS 8, July, 2010, Ljubljana, Slovenia. Voorburg: International Statistical Institute.

Reading, C., & Shaughnessy, J. M. (2004). Reasoning about variation. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 201–226). Dordrecht: Kluwer Academic.

Rouncefield, M. (1990). Classroom practicals using the binomial distribution. In *Proceedings of the third international conference on teaching statistics* (ICOTS 3). Retrieved from http://www.stat.auckland.ac.nz/~iase/publications.php.

Sánchez, E., & Landín, P. (2011). Fiabilidad de una jerarquía para evaluar el razonamiento probabilístico acerca de la distribución binomial. In M. Marín, G. Fernández, L. Blanco, & M. Palarea (Eds.), *Investigación en Educación Matemática XV*, Ciudad Real, España.

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.

Tarr, J. E., & Jones, G. A. (1997). A framework for assessing middle school students' thinking in conditional probability and independence. *Mathematics Education Research Journal*, *9*(1), 39–59.

Van Dooren, W., De Bock, D., Depaepe, F., Janssens, D., & Verschaffel, L. (2003). The illusion of linearity: expanding the evidence towards probabilistic reasoning. *Educational Studies in Mathematics*, *53*(2), 113–138.

Watson, J. M. (2006). *Statistical literacy at school (growth and goals)*. Mahwah: Erlbaum.

Watson, J. M., & Moritz, J. B. (1999a). The beginning of statistical inference: comparing two data sets. *Educational Studies in Mathematics*, *37*(2), 145–168.

Watson, J. M., & Moritz, J. B. (1999b). The development of the concept of average. *Focus on Learning Problems in Mathematics*, *21*(4), 15–39.

Children's Constructions of a Sample Space with Respect to the Law of Large Numbers

Efi Paparistodemou

Abstract The chapter describes how children use an expressive microworld to articulate ideas about how to make a game seem fair or not with the use of randomness. An open computer game was designed for children to express understanding of randomness as formal conjectures, so that they were able to examine the consequences of their understanding. The study investigates how 23 children, aged between $5\frac{6}{12}$ and 8 years, engaged in constructing a crucial part of a mechanism for a fair or not spatial lottery machine (microworld). In particular, the children tried to construct a fair game given a situation in which the key elements happened randomly. The children could select objects, determine their properties, and arrange their spatial layout in the machine. The study is based on task-based interviewing of children who were interacting with the computer game. The findings identify children's initial meanings for expressing stochastic phenomena and describe how the computer tool-based game helped shift children's attempts to understand randomness from looking for ways to control random behaviour, towards looking for ways to control events. Evidence is presented that the children constructed a set of situated abstractions (Noss 2001) for ideas such as the Law of Large Numbers. The computer game offered children the opportunity to make their own constructions of a 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.

1 Overview of the Problem

People take their chance in everyday life activities. For example, they take a chance of parking their car in the first available spot or moving on to find a place closer to their destination. Intuitive probability is something that is widely applied in everyday life, in making decisions and understanding social and natural phenomena.

E. Paparistodemou (✉)
Cyprus Pedagogical Institute, P.O. Box 12720, 2252 Nicosia, Cyprus
e-mail: e.paparistodemou@cytanet.com.cy

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_32,

Fischbein (1975) claims that intuitions develop as a kind of 'knowledge from experience', and that are used to take control over actions. Moreover, diSessa (1988) adds that intuitions consist of a number of fragments rather than one or even any small number of integrated structures one might call 'theories'. In a study of intuitions and fairness, Pratt and Noss (2002) showed how students older than the children of the present study made sense of dice-based situations. They illustrate how existing intuitions about fairness, often based on actual outcomes, are co-ordinated with new meanings, and they derive from interacting with a computational microworld. They also concluded that some of the meanings that their subjects developed were insufficient as ways of making sense of randomness, because this naturalistic meaning-making encouraged a data-oriented view of the world, where Kahneman and Tversky's (1982) misconceived heuristics can flourish. These naïve theories of stochastics are abstracted from experience with dice and other kinds of random generators, like cards, during informal games playing. It is here that technology has some real potential, as has been shown, for example, by Wilensky and Resnick (1999), Pratt (2000), Konold and Pollatsek (2002), Abrahamson (2009), demonstrating how the most obvious traps described by Tversky and Kahneman (1982) can, under the appropriate conditions, be circumvented.

Pratt and Noss (2002) describe the trajectory of students' thinking from an exclusive focus on short-term behaviour of stochastic processes to the emergence of new meanings which differed fundamentally from those that were held initially, and which emerged from the interaction with the computational system and the designed activity structures. These new meanings represented abstractions of long-term behaviour, a sense of an invariant relation that connected the number of trials to the specific configuration of the computational system. They reported children articulating new mental resources such as 'the more times you throw a die, the more even is the pie chart', in apparent recognition of what might be regarded as a situated version of the Law of Large Numbers. The authors referred to this statement as a situated abstraction. The current study adopted the notion of situated abstraction to refer to an action, taking place within the computational medium, typically consisting of a combination of manipulations of the spatial arrangements of balls as well as those actions are explained and described by the students.

The general theme that ran throughout this study is to explore the ways in which a specially designed computer game gave children the opportunity to develop and express probabilistic ideas. This study aims to describe and analyse how the tool-based game mediated the children's constructions of a sample space with respect to the Law of Large Numbers.[1]

[1]This study is based on aspects of my doctoral research (Paparistodemou 2004). I acknowledge Professor Richard Noss and Professor Dave Pratt for their help and support throughout this study.

2 Background

Descriptions of the chance of something happening are generally intended to downplay ignorance, while accounts of lack of knowledge point attention to the need to learn more. A quantitative description of the chance of an event occurring asserts implicitly that we have learnt all we can about it. This form of description invites us to stop asking questions and to start acting. Chance is linked with other terms like 'risk', 'accident', 'opportunity', 'randomness', 'possibility'. In the words of Gal (2005), randomness is a 'slippery' concept, which eludes a straightforward definition. However, among educators, there seems to be consensus that random phenomena are the ones that cannot be explored and explained through deterministic or causal means (Langrall and Mooney 2005). Falk et al. (1980) argue that probability is composed of two sub-concepts: chance and proportion. One has to be aware of the uncertain nature of a situation in order to apply the results of proportional computations. Obviously, the ability to calculate proportions as such does not necessarily signify understanding of probability. A realisation of uncertainty either in controlling or in predicting the outcome of an event is crucial.

Probability can be thought of as the mathematical approach to the quantification of chance, just as a ruler measures distances. The idea of risk reflects a fundamental yearning in humans, and perhaps all self-conscious beings, to know what will be; risk is a cultural concept employed in modern negotiations of everyday life as when, for example, somebody fastens his/her seatbelt before driving a car. The mathematical theory of risk provides a methodology for attempting to predict and control the future. It can also be defined as the projection of a degree of uncertainty about the future on the external world. The Royal Society (according to Heyman and Henriksen 1998) defined risk as the probability that a particular adverse event occurs during a stated period of time, or resulting from a particular challenge. Heyman and Henriksen (1998) argue that probability can arise from two sources: the randomness of events in the world, and ignorance. Based on this view, someone who chooses heads or tails in a coin toss has 0.5 probability of being correct because of the inherent randomness of coin tossing. On the other hand, a person who loses his/her way and chooses to turn left or right has a 0.5 probability of going in the right direction given a lack of knowledge of the geography of the area. In other words, uncertainty about the future, from this perspective, may arise out of either the randomness of the world or lack of knowledge. The distinction between chance and lack of knowledge turns not only on the complexity, relative to observers' meanings, of the future that is being predicted, but also on their pragmatic concerns.

To conclude, probability theory supplies information about the likelihood that data could have resulted from chance alone and it can be used to make decisions about uncertain outcomes. Understanding of the probability of an event, for the purposes of this study, is exhibited by the ability to identify and justify which of two or three events are most likely or least likely to occur (e.g. to understand the behaviour of an unfair die, which has one of six numbers twice). Many researchers have investigated young children's understanding of probability of an event. For example, Acredolo et al. (1989) note that children commit themselves to one of three strategies in comparing event probabilities: a numerator strategy in which they only examine the part that corresponds to the event, an incomplete denominator strategy in which they examine the part that corresponds to the complement of the event, and an integrating strategy in which they recognise the moderating effect that each part has on the other.

2.1 Understanding Sample Space

Consider an experiment whose outcome is not predictable with certainty. Although the outcome of the experiment will not be known in advance, let us suppose that the set of all possible outcomes is known. This set of all possible outcomes of an experiment is known as the *sample space* of the experiment and is denoted by S. A basic understanding of sample space is exhibited by the ability to identify the complete set of outcomes in a one-stage experiment. For example, if the outcome of an experiment consists in the determination of the sex of a newborn child, then $S = \{g, b\}$, where the outcome g means that the child is a girl and b that it is a boy. The *distribution* of a random variable is concerned with the way in which the probability of its taking a certain value, or a value within a certain interval, varies. More commonly, the distribution of a discrete random variable is given by its 'probability mass function' and that of a continuous random variable by its 'probability density function'.

I will also illustrate here the *Strong Law of Large Numbers*, as this is probably the best-known result in probability theory. It states that the average of a sequence of independent random variables having a common distribution will, with probability 1, converge to the mean of that distribution. The theorem of the Strong Law of Large Numbers says: Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having a finite mean $\mu = E[X_i]$. Then, with probability 1, $\frac{X_1+X_2+\cdots+X_n}{n} \to \mu$ as $n \to \infty$ (Ross 2002).

Borovcnik and Bentz (1991) suggest that symmetry played a key role in the history of probability. They claim that Laplace's attempt to define probability is characterised by an intimate intermixture of sample space (the mathematical part) and the idea of symmetry (the intuitive part). Although, in modern mathematics, the sample space is completely separated from the probability, which is a function defined on a specific class of subsets of sample space, the concept of sample space cannot fully be understood if it is not related to intuitions of symmetry. Moreover, Chernoff and Zazkis (2011) add to the literature of sample space the idea of a sample set. They define a sample set as any set of all possible outcomes, where the elements of this set do not need to be equiprobable. They refer to a sample set as learners' initial ways of reasoning, and build on it towards the idea of a sample space.

Piaget and Inhelder (1975) suggested that children before eight years of age are 'prelogical', but after eight years of age are able to identify all possible outcomes in a one-stage experiment. They found than children at the pre-operational level of development, which is common for children between 4 and 7 years old, are not able to differentiate between certainty (non-randomness) and uncertainty (randomness), because they lack particular intellectual operations, such as spatial–temporal a logical–arithmetical. Moreover, Jones et al. (1997) reported that significant numbers of children in elementary school (before 11 years of age) were not able to list the outcomes of a one-stage experiment. From the sample space one can identify whether an event is possible, and its possibility, impossibility or certainty (Jones et al. 1997). (The meanings of these terms are as follows: 'Possible' is equivalent to $0 < P(E) < 1$; 'Impossible' to $P(E) = 0$; and 'Certain' to $P(E) = 1$.)

3 The Design of the System

It was central to the design of the system, used in this study, that the children should not only see it as fun but also that such engagement would make it possible to observe thinking-in-change. Similarly, Stohl and Tarr's (2002) research focuses on how notions of formal stochastic inference can be fostered in middle school through the use of carefully designed tasks and open-ended software simulation tools. They analysed interactions between two sixth-grade students who used software tools to formulate and evaluate inferences. Their results suggest that young adolescents can develop powerful notions about inference when using simulation tools. They can recognize the importance of using larger samples in drawing valid inferences about probability, and use data displays to understand the bi-directional connection between theoretical and empirical probabilities.

A computer-game was designed in which children could express in action their own ideas about fairness and randomness. The basic idea was a game in which a 'space kid' moved upwards and downwards on a yellow line (the yellow line in Fig. 1 refers to the starting position of the space kid when the game starts), controlled by the position and movement of the balls in the lottery machine: the large square in the bottom right hand corner.

In practice, the children used a lottery machine rather than dice. In particular, a spatial lottery machine was designed. Children could place the balls in the lottery machine by themselves. A lottery machine, in which a small white ball moved in a straight line in a random direction, bounced and collided continually with a set of static blue and red balls controlled the movement of the space kid. By "continually" it's meant that the white ball kept colliding with the static balls and stopped only when the child ended the game manually or the space kid touched a mine. When the game was switched off, the white ball was static and placed in the centre of the lottery machine (the children could change its starting position). When the game started, the white ball moved randomly in a straight line until it collided with a coloured ball. Each time the white ball collided with a coloured ball or the walls of the lottery machine, it changed direction at a random angle and kept moving in a straight line. Each collision of the white ball with a blue ball (blue balls appear in the figures as the light grey) added one point to the 'blue score' and thus moved the space kid one step down the screen (blue is down). Conversely, each collision of

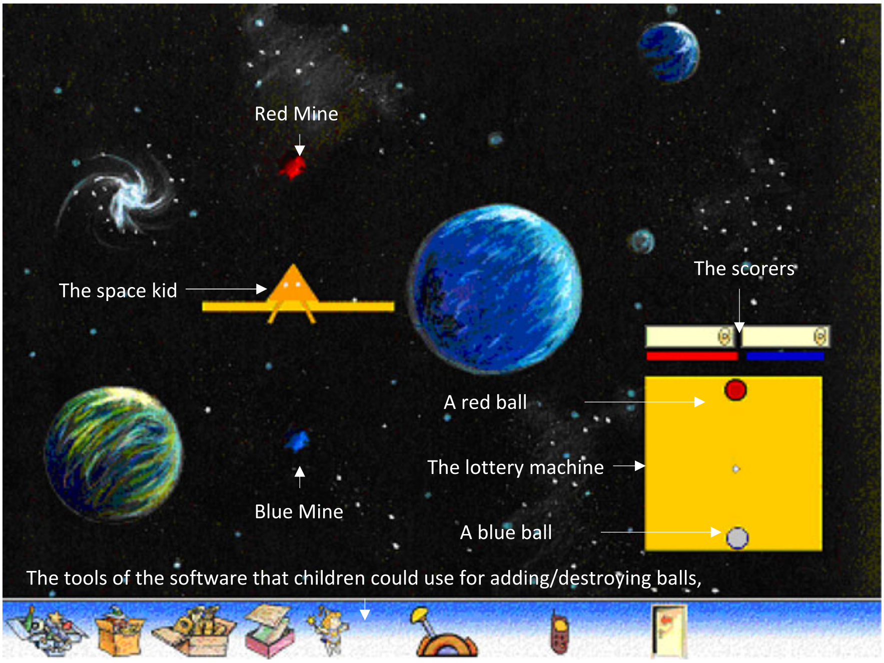


Fig. 1 A screenshot of the Space Kid game

the white ball with a red ball (red balls appear in the figures as the dark grey) added one point to the 'red score' and thus moved the space kid one step up the screen (red is up). Thus, the space kid's vertical location constantly indexed the number of blue and red collisions. For example, if the numbers of blue and red collisions were equal the space kid remained on the yellow line. Whilst individual collisions could be seen as single trials in a stochastic experiment, the totality of these movements gave an aggregated view of the long-term probability of the total events. Children could use the tools of the software to change and manipulate a number of aspects to construct their own lottery machine: the number, the size and/or the position of the balls. Moreover, the link between the objects of the game was visible through the 'rules' programmed onto the objects (for more details, see Paparistodemou et al. 2008).

The blue and red hits could be seen as an instantiation of the sample space. Thus, the lottery machine drove a game of chance, in which it became the visible engine for the generation of random events. By moving objects around inside the machine, the learner could stochastically or deterministically manipulate the outcome of the game according to his/her decision about how to manage the configuration of balls inside the lottery machine. The idea was to provide children a way to simultaneously see on screen the local and global representation of an event in their sample space. Several studies (for example, Wilensky and Resnick 1999; Ben-Zvi and Arcavi 2001; Konold and Pollatsek 2002; Rubin 2002) have illustrated the importance of linking local and global understanding with an aggregate view, and how complex students may find this process.

The representation of the lottery machine was geometrical/spatial, rather than in a textual format (for example, Pratt 2000), hidden altogether, or represented only in quantitative form (for example, Konold et al. 1993). Accordingly, the lottery machine was constructed so that its spatial arrangement might be seen to control global aggregated behaviour (see also Paparistodemou and Noss 2004).

4 Methodology

4.1 Context and Participants

The design of the game consisted of three main iterations. Twenty-three children, aged between $5\frac{6}{12}$ and 8 years old, participated in the final iteration reported here. The children worked with the software individually for a period lasting between 2 and 3 hours. They were described by their teachers as mid-socio-economic and mid-range of mathematical ability. During their interaction with the game, the children were guided through the activities with a series of verbally posed semi-structured questions, so as to be as unobtrusive as possible (although, inevitably, questioning children while they work does intrude into their activities and so the interventions formed part of the analysis). Interventions were used to validate and probe more deeply the thinking behind children's actions: The researcher acted as a participant observer. This stance indicates that as well as observing through participating in activities, the observer can ask the subjects to explain various aspects of what is going on Burgess (1984). The role of the researcher fell into probing interventions, experimental interventions and technical interventions (Paparistodemou 2004). Probing interventions aimed to make children's thinking transparent when it came to inferring the reasons that might lie behind their actions. Experimental interventions sought to make some change in the directions of the activity with possible implications for conceptual change. Technical interventions were made to give explanations about the software.

4.2 Data Collection and Data Analysis

All interactions and responses were videotaped, audio-recorded and transcribed. The children in the final iteration worked individually (previous iterations had demonstrated the extreme difficulty of making sense of an individual's understanding in paired activity). Children were asked to try out their ideas, describe what happened, give their explanation of why it had happened, and if it did not work as intended, what else they could do to reprogram it. The activities of the task aimed at providing open-ended opportunities for the construction of meanings, affording a wide range of opportunities to use the game for the children to express their ideas about chance, distribution, emerging pattern, and so on in action. The initial coding scheme for the data analysis was based on the preliminary interpretations of the literature review and on the aims of the study in tandem with the author's experience as an educator interacting with children of this age. The initial scheme was based on three broad provisional headings: Expressing-in-action Randomness and Fairness, Certain and Impossible Events, Probability of an Event. This study focuses on the children's expressions-in-action of chance and the law of the large numbers, which were included under the heading Expressing-in-action Randomness and Fairness.

5 Findings

The computer environment supported several ways in which the idea of mixture could be expressed. 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 judgements whether their own construction in the lottery machine was fair or not. Their constructions can be categorised into: (i) increasing the speed of the white ball (expressed by 7/23 children), (ii) adding more coloured balls (expressed by 16/23 children), (iii) adding more white balls (expressed by 21/23 children), (iv) making the size of the white ball(s) bigger (expressed by 5/23 children), and (v) leaving the game to work for a longer time (expressed by 23/23 children).

(i) 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 that allows you to change the size and the speed of an object] 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.

(ii) 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.

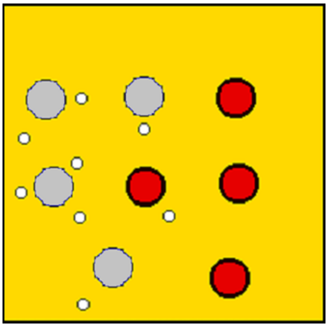


Fig. 2 Fiona's construction of fairness to get 'more points'

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 the importance of symmetry in his construction (Borovcnik and Bentz 1991) and 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.

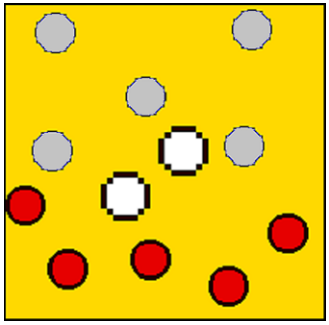
(iii) Adding more white balls

Adding more white balls was a strategy used by 19 out of 23 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 touch coloured balls more easily.

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! [wand is a tool in the game that allows her to copy objects]. Let's try it on (see Fig. 2).
She starts the game.

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 (Pratt and Noss

Fig. 3 Mathew's fair construction to get 'more points'



2002) 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 gave 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.

(iv) 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 Fig. 3.)
R: Now, we have 5 reds and 5 blues, how many points will they get?
Mathew: Many points. . . they will get equally many points.

Mathew made changes to get bigger 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.

(v) Leaving the game to work for a longer time

Another idea for getting bigger numbers in the game is not by making any changes in the mechanism of the game, but by waiting for the game to run longer. 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.

Demis (7 5/12 year-old boy) also did the same thing with the planets, giving his explanation as follows: 'Ah! …I have to move these two planets as well… I will put them down here and up there… not to touch them'. The idea was for the space kid not to touch them so quickly. Another child, Mathew, explained:

Mathew:	…But, it (the space kid) touches them. Shall I put them (the planets) a little farther away from it?
Researcher:	If you want to. Why?
M:	I will put them here….
R:	It (the space kid) won't get them there.
M:	Ok! I will put them here then. To take a while to get them. I will also put a fairy there, to watch them. I will put it here.
R:	Will our space kid get a planet now?
M:	Maybe if it goes too high.
	He starts the game again.
M:	It will take a *long time* to go to planet. It will move up and down…. They (the scorers) get equal points….

Mathew wanted to move the two planets in order for his game to work for a 'long time'. The two planets had been placed at the beginning, by the researcher, near the yellow line. Seeing this, Mathew realized that the space kid would touch them at once. So, he wanted to place the planets farther away, at the edges of the screen, in a position that it would be 'impossible' for the space kid to reach them. His strategy is to let his game run for a 'long time', so that the scores equalise.

6 Discussion and Didactical Implications

This paper aimed to explore the ways in which a specially designed computer game provided children the opportunity to develop and express probabilistic ideas. It might have been expected that children of this age would find it extremely difficult to construct a rich set of meanings for the challenging idea of randomness. The study indicates that probabilistic reasoning is influenced by the nature and structure of the particular task or problem situation (Langrall and Mooney 2005). Children have various cognitive resources for constructing randomness, beyond those that might be expected from classical experiments. A possible reason for this is that the tool offers them the opportunity to use these resources for random mixture in a two-dimensional continuum.

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 gave students situated abstractions (Noss 2001) 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 adds to Metz (1998) finding and it can be said that children by experiencing the game 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. The study demonstrated that students have robust intuitions for the Law of Large Numbers and that—given interactive learning environments—students can express these intuitions. This finding provides support to the general constructionist thesis (Papert 1996) that engagement in the building of some external, shareable, and personally meaningful product is conducive to mathematics learning.

There is no denying that in early age—and not only—chance and games are connected in the learning of probability. This study shows how a computer game and spatial representation introduce significant concepts of chance events, like the idea of Law of Large Numbers. Maybe spatial representation *in* the idea of chance has a role to play in teachers' instructional practices as they teach the topic of probability in their classes (Even and Kvatinsky 2010).

Young students bring intuitions to the classroom environment, but most of these are neglected because of the absence of teaching instruction. The findings of the study suggest that teaching probability can be based on the knowledge of students' intuitions, and an active learning environment can give students the opportunity to construct their own meanings by connecting new information to what they already know. In this chapter, I have described the evolutions in students' meanings for mathematical knowledge, which occurred while they built computational models of a sample space. The learner becomes involved in creating, as well as appropriating, artefacts that become part of the 'culture' of a learning system. I believe that a well-structured curriculum on probability could be introduced from the early levels of the elementary school. I suggest that probability should be introduced at this level for a variety of reasons: (a) The findings of this study suggest that intuitions on probability exist from an early age and therefore should be exploited before being abused; (b) Probability at this age can be connected with games and this can bring positive attitudes to pupils about mathematics and the connection to everyday life; (c) It helps one to understand and evaluate information in the world; (d) In later years, it is a prerequisite to enter many fields of study, and weak foundations of understanding are highly damaging. For these reasons, I believe that the subject of probability should not be neglected but should have a significant place in the elementary curriculum.

References

Abrahamson, D. (2009). Embodied design: constructing means for constructing meaning. *Educational Studies in Mathematics*, *70*, 27–47.

Acredolo, C., O'Connor, J., Banks, L., & Horobin, K. (1989). Children's ability to make probability estimates: skills revealed through application of Anderson's functional measurement methodology. *Child Development*, *60*, 933–945.

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

Borovcnik, M., & Bentz, H. (1991). Empirical research in understanding probability. In R. Kapadia & M. Borovcnik (Eds.), *Chance encounters: probability in education* (pp. 73–105). Dordrecht: Kluwer.

Burgess, R. (1984). *In the field*. New York: Routledge.

Chernoff, E., & Zazkis, E. (2011). From personal to conventional probabilities: from sample set to sample space. *Educational Studies in Mathematics*, *77*, 15–33.

diSessa, A. (1988). Knowledge in pieces. In G. Forman & P. Pufall (Eds.), *Constructivism in the computer age* (pp. 49–70). Hillsdale: Erlbaum.

Even, R., & Kvatinsky, R. (2010). What mathematics do teachers with contrasting teaching approaches address in probability lessons? *Educational Studies in Mathematics*, *74*, 207–222.

Falk, R., Falk, R., & Levin, I. (1980). A potential for learning probability in young children. *Educational Studies in Mathematics*, *11*, 181–204.

Fischbein, E. (1975). *The intuitive sources of probabilistic thinking in children*. London: Reidel.

Gal, I. (2005). Towards "Probability Literacy" for all citizens: building blocks and instructional dilemmas. In G. Jones (Ed.), *Exploring probability in school: challenge for learning and teaching* (pp. 39–63). New York: Springer.

Heyman, B., & Henriksen, M. (1998). *Probability and health risks. Risk, health and health care* (pp. 65–105). London: Arnold.

Jones, G., Langrall, C., Thornton, C., & Mogill, A. (1997). A framework for assessing and nurturing young children's thinking in probability. *Educational Studies in Mathematics*, *32*, 101–125.

Kahneman, D., & Tversky, A. (1982). On the study of statistical intuitions. *Cognition*, *11*, 123–141.

Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education*, *33*(4), 259–289.

Konold, C., Pollatsek, A., Well, A., Lohmeier, J., & Lipson, A. (1993). Inconsistencies in probabilistic reasoning of novices. *Journal for Research in Mathematics Education*, *24*, 392–414.

Langrall, C., & Mooney, E. (2005). Characteristics of elementary school students' probabilistic reasoning. In G. Jones (Ed.), *Exploring probability in school: challenge for learning and teaching* (pp. 95–119). New York: Springer.

Metz, K. (1998). Emergent understanding and attribution of randomness: comparative analysis of reasoning of primary grade children and undergraduates. *Cognition and Instruction*, *16*(3), 285–365.

Noss, R. (2001). For a learnable mathematics in the digital culture. *Educational Studies in Mathematics*, *48*, 21–46.

Papert, S. (1996). An exploration in the space of mathematics educations. *International Journal of Computers for Mathematical Learning*, *1*, 95–123.

Paparistodemou, E. (2004). *Children's expressions of randomness: constructing probabilistic ideas in an open computer game*. PhD Dissertation, Institute of Education, University of London.

Paparistodemou, E., & Noss, R. (2004). Designing for local and global meanings of randomness. In *Proceedings of the twenty eighth annual conference of the international group for the psychology of mathematics education*, Bergen, Norway (Vol. 3, pp. 497–504).

Paparistodemou, E., Noss, R., & Pratt, D. (2008). The interplay between fairness and randomness in a spatial computer game. *International Journal of Computers for Mathematical Learning*, *13*(2), 89–110.

Piaget, J., & Inhelder, B. (1975). *The origin of the idea of chance in children*. London: Routledge & Kegan Paul.
Pratt, D. (2000). Making sense of the total of two dice. *Journal for Research in Mathematics Education*, *31*(5), 602–625.
Pratt, D., & Noss, R. (2002). The micro-evolution of mathematical knowledge: the case of randomness. *The Journal of the Learning Sciences*, *11*(4), 453–488.
Ross, S. (2002). *A first course in probability* (6th ed.). Upper Saddle River: Prentice Hall.
Rubin, A. (2002). Interactive visualizations of statistical relationships: what do we gain? In B. Phillips (Ed.), *Proceedings of the sixth international conference on teaching of statistics* (ICOTS 6), Cape Town. Voorburg: International Statistical Institute.
Stohl, H., & Tarr, J. (2002). Developing notions of inference using probability simulations tools. *The Journal of Mathematical Behavior*, *21*(3), 319–337.
Wilensky, U., & Resnick, M. (1999). Thinking in levels: a dynamic systems perspective to making sense of the world. *Journal of Science Education and Technology*, *8*(1), 3–18.

Researching Conditional Probability Problem Solving

M. Pedro Huerta

Abstract The chapter is organized into two parts. In the first one, the main protagonist is the conditional probability problem. We show a theoretical study about conditional probability problems, identifying a particular family of problems we call ternary problems of conditional probability. We define the notions of Level, Category and Type of a problem in order to classify them into sub-families and in order to study them better. We also offer a tool we call trinomial graph that functions as a generative model for this family of problems. We show the syntax of the model that allows researchers and teachers to translate a problem in terms of the trinomial graphs language, and the consequences of this translation.

In the second part, there are two main related protagonists: ternary problems of conditional probability and students solving them. Thus, the students' probabilistic thinking is observed in a broader problem-solving context, in relation to the task variables of problems: structure, context and data format variables. We report some of the results of our investigation into students' behaviours, showing how these depend in any manner on those task variables.

1 Introduction

An alternative perspective in which probabilistic thinking could also be considered is offered in this article. This alternative consists in placing probabilistic thinking in a problem solving framework. This article, although implicitly, also formulates an invitation to consider probability problem solving as a research field for the probabilistic educators and researchers community, still not systematically tackled, as we can check, for example, in Jones and Thornton (2005).

Basically, researching into probability problem solving should involve three elements, individually and in relation to each other: problems, students solving these problems and teachers teaching students to solve them. About the problems themselves we have very little information. We do not know what problems students

M.P. Huerta (✉)
Departament de Didàctica de la Matemàtica, University of Valencia, Avenida de Tarongers, 4, 46022, Valencia, Spain
e-mail: huerta@uv.es

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_33,

have to solve, what characteristics problems possess that might have an influence on students' behaviour and which make them so difficult to solve. We have not yet considered in which direction teaching models should be addressed so that they could help students to become competent and probabilistically literate in the actual uses of the probability. It is hoped that in this chapter researchers and teachers can find hints as to in which direction future approaches to teaching conditional probability could be based, such as, for example, exploring context of uses by means of solving problems (Carles and Huerta 2007; Huerta 2009).

In this article, in particular, we talk about conditional probability problems, offering a systematic method to study them. We show a theoretical study into those problems; identifying a particular family of problems we call *ternary problems* of conditional probability, and we classify them into families and sub-families to study them better. We also offer a tool we call *trinomial graph* that functions as a generative model (in the sense of Fischbein 1975) for this family of problems and that we consider would be useful for teachers and researchers. We show the syntax of the model that allows the translation of a problem in terms of trinomial graphs, and the consequences of this translation. Among others, one of their uses is shown in this part of the article: how these problems can be analysed before they become a task for students, depending on the task variables (in the sense of Kulm 1979) researchers and teachers want to take into account in their works.

At the end of the article, there are two main related protagonists: ternary problems of conditional probability and students solving them. As a result, the students' probabilistic thinking is observed in a broader problem solving context, in relation to the task variables: structure, context and data format. We report some of the results of our investigation into students' behaviour, showing how it depends in any manner on those task variables.

2 The Notion of Ternary Problems of Conditional Probability in This Work

We say that a school task is a *probability problem* if in its formulation at least one quantity can be interpreted as a probability, either known or unknown. This quantity can be expressed in the text of the problem in the normal way, such as frequencies, rates, percentages and probabilities. In general, as a school task, these problems are word-problems and hence they are formulated in some particular context. As problems, tasks always allow a reading in terms of probabilities and as a consequence it is possible to re-formulate them using the formal language of the probabilities.

We say that a probability problem is a *conditional probability problem* if in its formulation at least one of the quantities explicitly mentioned in the problem could be interpreted as a conditional probability. In general, a problem will be thought of and formulated in a mathematical context. This problem, formulated as a task, will be contextualized but not necessarily mathematically. This formulation will be affected by the characteristics of the context: language, specific terminology, data format for the numerical data, different expressions for the quantities and their relationships, and other. Thus, these conditional probability problems are word-problems in order to distinguish them from those formulated using the formal language of the probabilities.

A *ternary problem of conditional probability* is a conditional probability problem formulated with three known quantities and one unknown quantity, to be solved for, and verifying that all probabilities in the problem are connected by ternary relationships, that is to say, additive or multiplicative relationships between the three quantities (Huerta 2009).

Examples of ternary problems of conditional probability are well known problems, such as The Disease Problem and The Taxicab Problem, which can be found in Tversky and Kahneman (1983, pp. 154 and 156) and in Shaughnessy (1992, p. 471).

We define *quantity* in a conditional probability problem as a triple (x, S, F), with x being a number; S the description of the number and F the data format for the number. This x may be a natural number (or frequency), a relative number (rate or percentage, r or %) or a probability (number in the interval between 0 and 1, p), so S may be a sentence, a set or an event in a σ-algebra of events, respectively.

Finally, ternary problems of conditional probability can be modelled with two basics events A and B, their complementary events $\overline{A}$ and $\overline{B}$; 16 quantities, 4 of them referring to marginal probabilities, 4 to joint or intersection probabilities and, finally, 8 referring to conditional probabilities; and 18 ternary relationships, as follows:

- 10 additive relationships: 6 of them complementary relationships, as in $p(A) + p(\overline{A}) = 1$ and 4 referring to the total probability, like $p(A) = p(A \cap B) + p(A \cap \overline{B})$, and
- 8 multiplicative relationships, as many as there are conditional probabilities considered, like $p(A) \times P_A(B) = p(A \cap B)$.

3 Task Variables for the Ternary Problems of Conditional Probability

When we take a careful look at the task to be solved by students, we can observe how these problems have some characteristics that one can guess are going to be influential factors in students' solutions. That is to say, these characteristics could be influential factors, leading to the *difficulties* in the problems. As an example, let us consider the following problems:

Boys and girls

The fourth class of a secondary school has 30 pupils, boys and girls. Among these pupils, there are 16 who wear glasses; there are 17 girls and 10 girls who do not wear glasses. Among the boys of the class, what percentage of them wears glasses?

AIDS

A population at risk of suffering from AIDS undergoes a test to diagnose whether it has it or not. The test result is positive or negative in each case. The probability that one person at risk has AIDS is 0.57 and the probability that one of them has a positive result is 0.47. One knows also that there is a probability of 0.23 that a person suffers from AIDS and gets a negative test result. For the persons who do not have AIDS, what is the probability that they have a positive test result?

Antibiotic

A population suffers from a skin infection. Some of them have been treated with an antibiotic and some not. 53 % of the persons have been cured, 42 % have been treated with an antibiotic and 7 % have been treated with the antibiotic and have not been cured. Among the persons who have not been treated with the antibiotic, what is the percentage of those that have been cured?

If we make a reading of these problems in the mathematical context, then all three could be translated to this problem: given $p(A \cap \overline{B})$, $p(A)$, $p(B)$, calculate $P_{\overline{A}}(B)$. So, in some sense, we could say that all three problems are isomorphic. But, it is possible that a quick reading of the problems allows us, or solvers, to appreciate that a problem is more difficult than another. Appreciating these difficulties is not far from seeing the problem by means we call characteristics of the problem or task variables in the sense of Kulm (1979, p. 1).

The task variables described by Kulm are adapted and particularized for the ternary problems of conditional probability. Thus, for the purpose of this work, the task variables we will refer to throughout the article are the following: context variables, structure variables and format variables. Another task variable, the content variable, has been already fixed from the beginning of the article: the conditional probability.

Finally, just as in Kulm (1979, p. 8) the possibility is mentioned that the task variables are influential factors in the process of resolution of a mathematical problem, and exerting this influence, to a greater or lesser degree, on the phases of problem resolution (Puig and Cerdán 1988, adapted from Polya 1957). We can appreciate in the next diagram (Fig. 1) where and how task variables might have an influence on the phases of resolution of the ternary problems of conditional probability. To study these influences should be matter for future research.

It is reasonable to think that most of students' difficulties in solving these problems could be found in the relationships shown in the diagram above (Fig. 1). For example, the correct identification of the mentioned quantities in the text of the problem, both known and unknown, in every three components, allows students to understand the problem, and the correct choice of the accurate relationships between known and unknown quantities will produce new intermediate quantities in looking for the solution of the problem.

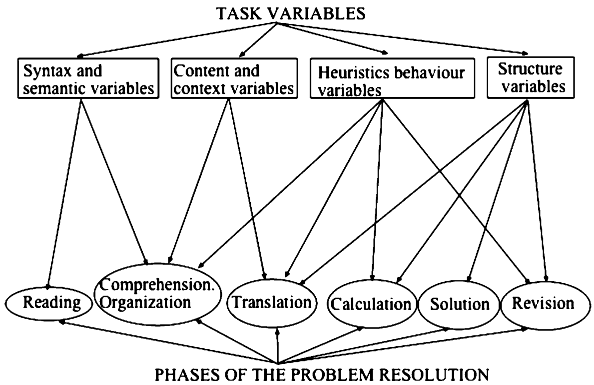


Fig. 1 Diagram relating task variables and phases of the problem resolution

4 The Trinomial Graph as a Generative Model for the *World* of Ternary Problems of Conditional Probability

From previous works by Cerdán (2007) about the family of arithmetical–algebraic problems, Cerdán and Huerta (2007) built a meta-language with which to model the family of ternary problems of conditional probability. Based in this meta-language, the result of its application to the set of quantities and relationships between quantities that characterizes problems from this family gives us the *trinomial graph* of the whole *world* of ternary problems of conditional probability (Huerta 2009). This meta-language is useful for doing analytical readings of problems, that is to say, readings in which users pay attention only to quantities, both known and unknown, and to the relationships between quantities.

Because it is a language, it uses a set of signs, rules and conventions that allow users to represent every ternary problem and to sketch as many resolutions of the problems as one can consider both theoretical resolutions and students' resolutions. But at the same time, given a problem on the graph and a context in which to formulate the problem, the translation from the meta-language of the graph into natural language allows teachers and researchers to turn a problem in the graph into a task for students.

4.1 Signs, Rules and Conventions

Let us consider the following version of the Taxicab problem:

A cab was involved in a hit and run accident at night. There are two cab companies that operate in the city, a Blue cab company, and a Green cab company. You know that:

(a) 85 % of the cabs in the city are Green (15 % are Blue, implicit data);
(b) A witness identified the cab as Blue.

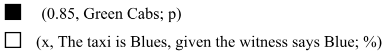


Fig. 2 Vertices or nodes representing quantities in the Taxicab problem

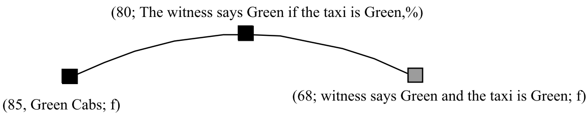


Fig. 3 Multiplicative oriented trinomial edge in the Taxicab problem

The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colours 80 % of the time (and failed 20 % of the time, implicit data). What is the probability that the cab involved in the accident was Blue rather than Green?

A vertex or node in the graph will represent a quantity in the text of the problem, either known or unknown. In a graph, quantities are represented by means of small squares labelled with a letter or a sentence that describes the number of the quantity. If this quantity is known, then we represent it as a dark vertex, if not then the vertex is white (Fig. 2). Next to each vertex we can write the quantity it is representing, as in the example:

We can simplify the written representation of quantities avoiding the third component of the quantity if there is no ambiguity or discordance between the three components of the triple. On the other hand, there is only one vertex in the graph, which we represent with a small dark circle. It represents some of the following known quantities in every problem: (1, probability space; p) or (100, measure of the sample space; %) or (n, size of the sample space, f) (see following comments in Fig. 6). So, depending on this vertex, all quantities in the problem will be homogeneous with it in numerical data format.

Because all relationships in ternary problems are ternary, they are represented by means of edges we call *trinomial edges*. As the relationships may be additive or multiplicative, these properties will be represented in the graph by means of dot lines or solid lines, respectively, connecting three vertices or nodes. Moreover, given that every relationship has a specific way of being read, then the trinomial edges incorporate this way of reading from the left to the right (Fig. 3). The following are examples of this:

(a) Multiplicative relationship

(85; *Green Cabs*; f) *x* (80 %; *Witness says Green, if the taxi is Green*, %) = (68, *Witness says Green and the taxi is Green*; f).

Notice that in Fig. 3 one vertex is coloured grey. This means that this quantity has been calculated by means of the relationships, given the other two.

(b) Additive relationships

(85, *Green Cabs,* f) = (17, *The witness says Green and the taxi is Blue,* f) + (68, *The witness says Green and the taxi is Green,* f).

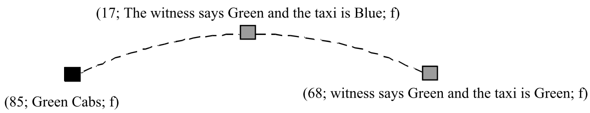


Fig. 4 Additive oriented trinomial edge in the Taxicab problem

In Figs. 3 and 4, we have represented the quantities whose numerical data formats are frequencies or percentages. If there is a reading of the problem in terms of probabilities, then quantities and relationships are represented in the same way but the expressions of the quantities next to each vertex are changed (Fig. 5).

Depending on numerical data format, the relationships of complementarity may converge to n, size of the sample space if quantities are expressed in frequencies, to 100 if percentages or to 1 if probabilities are considered. The next edges represent these relationships for every case.

It is not usual that the quantities referred to 1, 100 or 1000 are explicitly mentioned in the text of problems. In fact, they are known which is the reason why we represent them as a dark vertex, but because they are not explicitly mentioned in the problem we usually represent them as small circles and not squares.

4.2 A Problem Resolution Viewed as a Net of Enchained Edges

The Taxicab problem may be also formulated as follows: Known (0.85; G; p); (0.80; WB; P_B); (0.80; WG; P_G), calculate (x; B; P_{WB}).

The set of additive and multiplicative relationships, all quantities involved with their types and a resolution of the problem are represented in Fig. 7 as a net of enchained edges. We call this net the *graph of a theoretical resolution* of the problem,

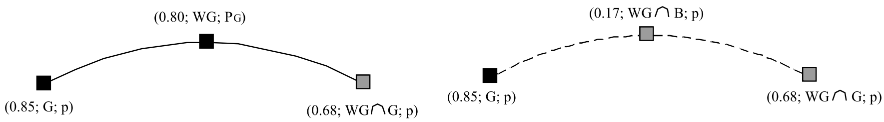


Fig. 5 Multiplicative (*left*) and additive (*right*) oriented trinomial edge in the Taxicab problem

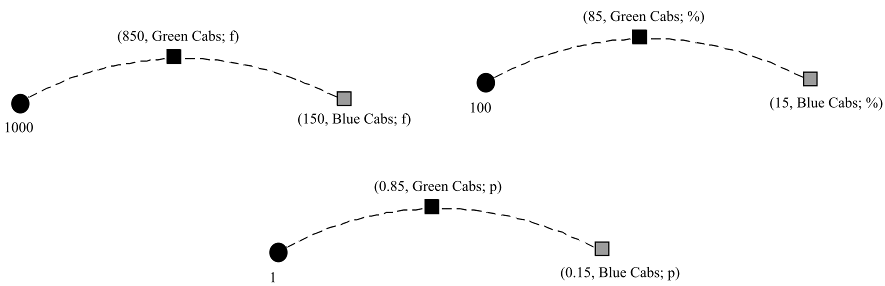
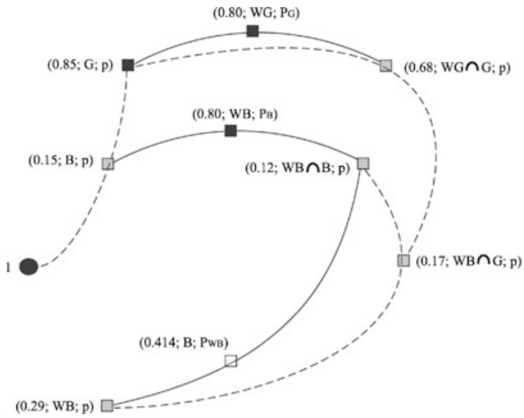


Fig. 6 Additive edges representing possible relationships of complementarity

Fig. 7 A trinomial graph of one resolution of the Taxicab problem



and it represents the least number of quantities and relationships that allow finding the answer to the question.

This graph (Fig. 7) gives us information about particularities of the resolution it represents. Thus, we can see that this resolution uses three additive and three multiplicative relationships for calculating the five required intermediate quantities (the grey squares). This set of relationships and intermediate quantities can give us a first approximation to the complexity of this solution of the problem. This complexity is dependent on the given known quantities and the quantity is asked for (the white square in Fig. 7). Graphs of other problems' solutions may differ in the number of calculated intermediate quantities and used relationships. Subsequently, the solutions of the problems become more complex but less efficient.

4.3 The Trinomial Graph of the World of Ternary Problems of Conditional Probability

The graph in Fig. 7 represents one theoretical solution of a ternary problem of conditional probability, but is not the only one. The task we set up for this problem could be anticipated for every possible resolution of every ternary problem.

Of course, we already commented that the particular world of the ternary problems of conditional probability can be modelled by 16 quantities and 18 relationships. In Fig. 7, we have also seen how a resolution of a ternary problem uses some of these quantities and a net of (additive and multiplicative) ternary relationships among the 18 possible relationships. So, if we represent all quantities and relationships together, we will obtain the graph in Fig. 8 showing us these quantities and the net of relationships.

We call the graph in Fig. 8 the *Trinomial Graph of the world of ternary problems* of *conditional probability*. The graph is thus named because every ternary problem of conditional probability can be formulated there by means of three appropriated dark vertexes and the corresponding white one (as the question which has to be

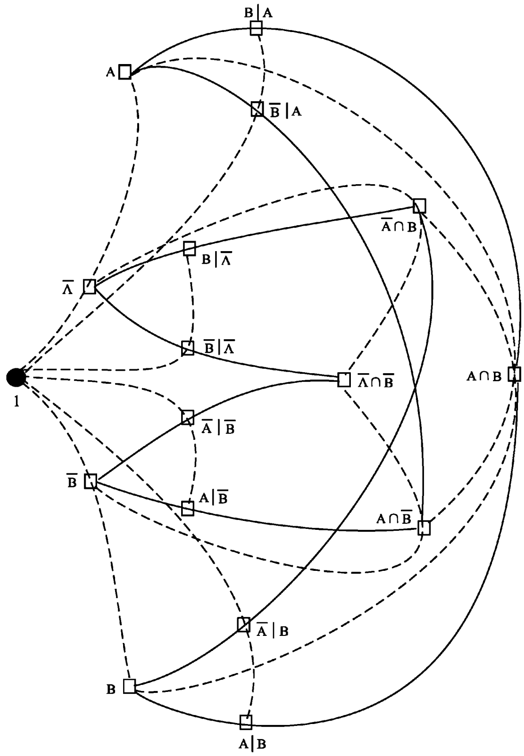


Fig. 8 The *Trinomial Graph* of the ternary problems of conditional probability *world*

solved) and, furthermore, every successful resolution of a problem has a place on the graph.

In order to make the representing of quantities as simple as possible, we have taken some decisions that need clarification. One is the decision to organize quantities in a symmetrical way in order to appreciate the symmetry of the events and their operations. The second one has to do with the representation of the quantities referring to conditional probabilities. We have opted to represent by $A|B$ everything related to in this article which has been expressed by $P_B(A)$; that is to say, the conditional probability of A given B or the probability of A conditioned by B. Users of this graph may adopt other representations of the quantities but always indicating in which sense and with which meaning they are using their signs for representation.

With the graph in Fig. 8, we have modelled a particular world of problems. As a model, the graph is a generative model, in the sense of Fischbein (1975, p. 110), especially useful for teachers and researches in probability problem solving, because:

- "... with a limited number of elements and rules for the combination, it can correctly represent an unlimited number of different situations".
- It is a heuristic model, "... by itself leading to solutions which are valid for the original (problem) as a result of the genuine isomorphism between the two realities involved" (the graph and the original problem).

- It is generative, in the sense that "it is capable of self-reproduction, in that its coding is sufficiently general for it to be able to suggest new models", for example, ternary problems involving $p(A \cup B)$ quantities.

4.4 The Algorithm for Solving Problems in the Trinomial Graph. Theoretical Readings of a Problem

Given that an edge may have all three vertexes white, two white and one dark, or one white and two dark, we define *entrance to the graph* to every edge with two dark nodes. Because of the additive or multiplicative character of an edge, given an edge with two dark nodes, the third is easily calculated. This calculated node, in its turn, might provide new entrances to the graph. Applying this repeatedly and given that the question is a white node; the problem will be solved when this white node is a dark one. We call to this algorithm the *edges destruction algorithm*. In Huerta (2009), an electronic demonstration of this algorithm can be seen.

Furthermore, if we use the analysis and synthesis method, that is to say, the method of problem solving that starting from the unknown quantity determines the required intermediate quantities strictly necessary to reduce the calculation of the unknown quantity to a calculation among known quantities in the text of the problem, then we have at our disposal what we call *theoretical readings of problems*. These theoretical readings are analytical, in the sense that they only pay attention to quantities and relationships between quantities. The next graph (Fig. 9) represents a theoretical reading of the Taxicab problem.

In the graph, we see known quantities (dark nodes), calculated quantities (grey nodes) and unknown quantities (white nodes). We also can appreciate the relationships between quantities that have been considered in the resolution of the problem (stressed lines).

With $p(A|B)$ being the question in the problem, that is to say, $P_{WB}(B)$, this is the synthesis from the graph:

$$p(\overline{A}|B) = \frac{p(\overline{A}) \times [1 - p(\overline{B}|\overline{A})]}{p(\overline{A}) \times [1 - p(\overline{B}|\overline{A})] + [1 - p(\overline{A})] \times p(B|A)}.$$

That is to say, for the symbols we use in the problem:

$$P_{WB}(B) = \frac{[1 - p(G)] \times P_B(WB)}{[1 - p(G)] \times P_B(WB) + [p(G) - p(G) \times P_G(WG)]}.$$

As we can see in the latter expression, the calculation of the quantity in question has been reduced to a calculation among known quantities, all of them in the text of the problem. Thus, in the graph (Fig. 9), starting from the dark nodes and applying the algorithm of destruction of edges, the white node representing the question has been darkened (calculated). If it is possible to do a reading like this in a conditional probability problem, we say that this problem has an analytical reading that is arithmetical, or simply it has an *arithmetical reading*, and subsequently, we say that the problem has an associated theoretical graph that is arithmetical.

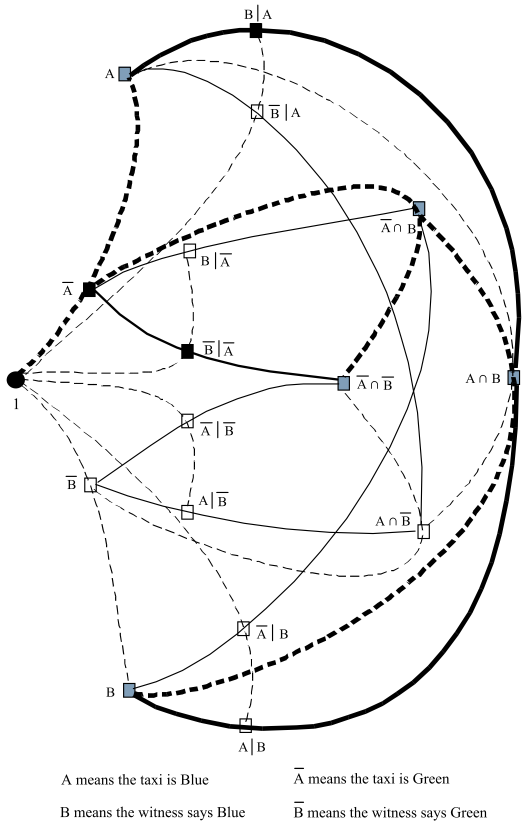


Fig. 9 An analytical reading of the Taxicab problem in a trinomial graph

But not all ternary problems of conditional probability have arithmetical readings. For example, let us consider the following problem:

A medical research lab proposes a screening test for a disease. To try out this test, it is given to a number of persons, a percentage of whom are known to have the disease. A positive test indicates the disease and a negative test indicates no disease. Unfortunately, such a medical test can produce errors because of:

1. Sensitivity to the disease: For the people who do have the disease, this screening test is positive in the 96.7 % of the cases.
2. Specificity for the no disease: For the people who do not have the disease, this screening test is negative in the 75 % of the cases.

On the other hand, this lab assures that a positive result is quite feasible because among people who have a positive result in the test 95 % do have the disease.

Suppose the test is given to a person whose disease status is unknown. What is the probability that he/she has the disease?

In Carles and Huerta (2007), we show a trinomial graph of the world of ternary problems in the context of the diagnostic test in Health Sciences. Let

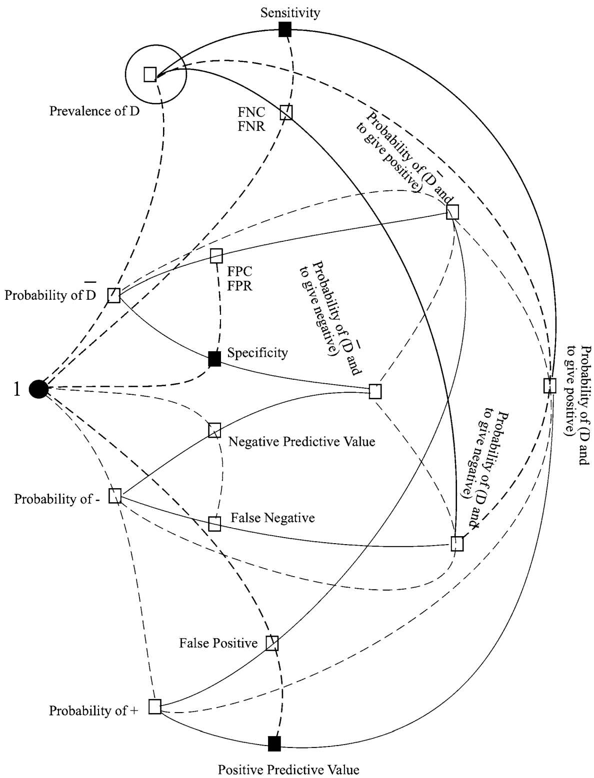


Fig. 10 Graph with the quantities in the medical diagnosis task. The question has been circled

us plot the known quantities in the problem onto the graph. This is the result (Fig. 10).

We only have three entrances to the graph. They represent the complementary relationships of the known quantities in the problem. If we complete them, then the process of solving the problem is as we can see in the next graph (Fig. 11).

In the graph in Fig. 11, we can see how the calculated intermediate quantities, the false positive coefficient (FPC), the false negative coefficient (FNC) and the False Positive, shape a graph without new entrances to it. A graph like this is not an arithmetical one. Therefore, we say that the problem has a theoretical reading that is algebraic, or simply that it has an *algebraic reading*, and hence that the problem has an associated theoretical graph that is algebraic.

With the objective of proceeding to search for the solution of the problem in the graph, it is necessary to consider a white node as if it were dark; this is to say, as if it were known. This node may be any that allows solvers to have a new entrance to the graph. Thus, if we assume that the prevalence of D is x, the graph has at least one entrance. Acting by means of the algorithm with the new quantity (x, prevalence of D, p) as if it were known, every new intermediate quantity will be an algebraic expression in x, as we can see in the next graph (Fig. 12).

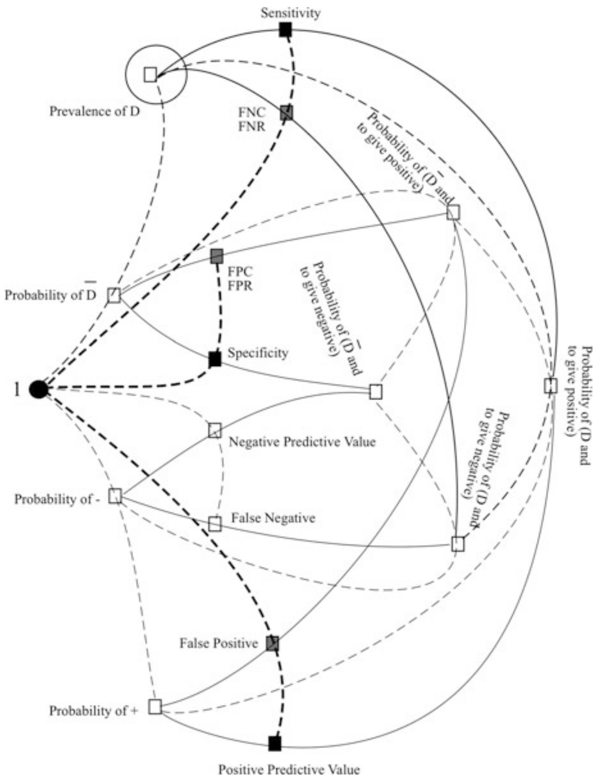


Fig. 11 State of the graph after applying the algorithm to the known quantities

At least one of the intermediated quantities can be calculated in more than one way. In our case, the quantity (z, probability of $+$, p) has been calculated by means of two edges:

– An additive edge
(z, probability of $+$, p) $= [0.25(1-x)$; probability of ($\overline{D}$ and to give positive); $p]$ $+$ $[0.967x$; probability of (D and to give positive); $p]$, giving: $z = 0.25(1-x) + 0.967x$.
– A multiplicative edge
(z, probability of $+$, p) $\times$ (0.95, Positive predictive value; p) $=$ $[0.967x$; probability of (D and to give positive); $p]$, giving this time: $z = \frac{0.967x}{0.95}$.

Then, necessarily, an equation can be formulated as follows:

$\frac{0.967x}{0.95} = 0.25(1-x) + 0.967x$, giving the value of $x = 0.83$, that is to say, the Prevalence of D is 83 % as the answer to the question.

By means of two examples, we have shown how ternary problems of conditional probability may have theoretical analytical readings that may be either arithmetical or algebraic. This does not mean that solvers necessarily do similar readings. Of course, solvers may do readings of a problem with theoretical arithmetical readings, which are either arithmetical or algebraic, as they may do with a problem with

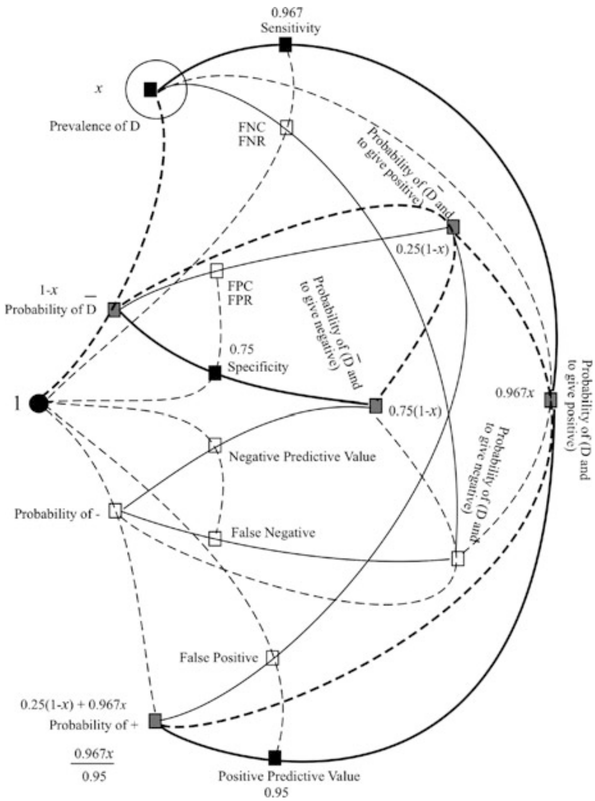


Fig. 12 Graph representing an algebraic resolution of the medical diagnostic task

theoretical algebraic readings. The importance of previous knowledge, for either teachers or researchers, of theoretical readings of problems in order to convert them into tasks for students is crucial in order to obtain knowledge about possible complexities and difficulties of the task students are going to solve. Students' behaviour may also depend on the kind of the theoretical reading of problems.

5 Classification of the Ternary Problems of Conditional Probability

Remember that, structurally speaking, a ternary problem of conditional probability can be identified by considering only three known quantities and one unknown asked quantity. In the language of the trinomial graphs, by means of three dark nodes and one white node that is required to be darkened (calculated). Because these four quantities have to necessarily belong to the set of marginal, joint or conjunction and conditional quantities involved in every ternary problem, an easy combinatorial analysis may allow us to identify the quantities that have to be considered, and with the help of the trinomial graph (Fig. 8) to decide to whether or not a ternary problem of conditional probability could be formulated with these four quantities.

Table 1 Classification of ternary problems of conditional probability into families and subfamilies

	L_0			L_1			L_2			L_3		
C_0	C_0T_1	$\varnothing$	$\varnothing$	C_0T_1	C_0T_2	C_0T_3	C_0T_1	C_0T_2	C_0T_3	C_0T_1	C_0T_2	C_0T_3
C_1	C_1T_1	$\varnothing$	$\varnothing$	C_1T_1	C_1T_2	C_1T_3	C_1T_1	C_1T_2	C_1T_3	$\varnothing$	$\varnothing$	$\varnothing$
C_2	C_2T_1	$\varnothing$	$\varnothing$	C_2T_1	$\varnothing$	C_2T_3	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$

Lonjedo (2007) carries out this task. As a result of her work, ternary problems have been classified into families and subfamilies of problems. This classification considers three structural variables in the problems, as the following describes:

- L_i, called the *level of a problem*. This level refers to the number of quantities that are formulated as conditional quantities (probabilities) in the text of the problems. There are four possible values for a level: 0, 1, 2, and 3. This means that a problem may be formulated giving a number of conditional probabilities between 0 and 3. The Taxicab problem in the Sect. 4.3 is an L_2-problem and the one in Fig. 10 is an L_3-problem.
- C_j, called the *category of a problem*. This variable is related to the level of the problem. Therefore, the level and the category of a problem must be taken together. Its values indicate the number of marginal quantities that are known in problems. In this way, and because these probabilities refer to the two basic events, problems may inform about the two events, about one of them or about neither. So, the possible values of this variable are: 0, 1 and 2. It is necessary to recall that we are referring to quantities explicitly mentioned in the text of the problem.

Given a problem, if we know the level and category of the problem, we automatically get information about the known joint quantities (probabilities). This is because of the total number of known quantities must be 3. The fourth quantity, the unknown quantity, is identified by means of the following variable:

- T_k, called the *type of the problem*. If the asked quantity is a conditional probability, we say the problem is a T_1-problem; if a marginal then T_2; and if a joint probability then of T_3 type.

In conclusion, we can say that every ternary problem of conditional probability can be structurally identified by the triple (L_i, C_j, T_k). This triple informs us of the data structure of the problem and perhaps about its potential difficulty. The Taxicab problem is an (L_2, C_1, T_1)-problem. With a few data we can easily describe the structure of data in a problem.

As a result of the combinatorial analysis with these variables, Lonjedo (2007) and Huerta (2009) report a classification of the problems, as we can see in Table 1. The *world* of problems has been organized into four L-families. Every L-family has been divided into subfamilies according to categories and types. Therefore, we can identify a total of four families and 20 subfamilies (Table 1).

In every case, the symbol $\varnothing$ means that there are no ternary problems described by the triple corresponding to the cell in which $\varnothing$ is. This does not necessarily mean that one cannot formulate problems with data structure that ∅ denies. We only want to express that the problems are neither conditional probability problems nor ternary problems.

Thus, for example, the family of problems with level L_0 can only be divided into three subfamilies, all of them with type T_1. Any other subfamily with types T_2 or T_3 will be a probability problem but not a conditional probability problem. In the case of the L_3-family of problems, all three known quantities are conditional probabilities. Consequently, no more known quantities are required to formulate problems. If so then problems will be extra-dimensioned and at least one known quantity could be removed from the text of problems, meaning that the new problem could be classified into other families if the one removed is a conditional probability; or into some of the subfamilies in the L_3-family if the one removed is a marginal or an intersection probability.

In the case of the L_1-family, the reason for which one subfamily is left is due to the meaning of the category of a problem. This subfamily is described by the triple (L_1, C_2, T_2), that is to say, any problem from this subfamily would be formulated by means of one conditional and two marginal probabilities. The question is one marginal probability. In a problem like this, nothing has to be done because the answer to the question in the problem is the complementary of one of the two known marginal probabilities. In fact, the question is not an unknown quantity, as it must be in every problem, it is already known in this case because the text of the problem has already given information about it, albeit implicitly.

Finally, problems from the L_2C_2-subfamilies, in all three types, are extra-dimensioned. So, we do not consider that they could constitute a family of problems. As extra-dimensioned problems, they could be reformulated without one of the quantities, thus becoming a new problem that could be classified into another subfamily depending on the quantity finally removed.

6 The Formulation of Tasks

Up to now, we have talked about problems, their elements and characteristics, their classification and their structural complexities. We have also talked about task variables as variables that we need to take into account, both for teaching and researching on conditional probability problem solving, because of their possible influence on the process of solving the problems. Now it is time to talk about tasks, as the result of the conversion of problems, as mathematical objects, into school-problems or tasks, with the aim that students in a mathematics class or subjects participating in research solve them. Obviously, these tasks have to respond to teaching or research objectives.

Let us imagine we want to teach or research problems from the L_0-family. It must be recalled that this means that in the text of the task no information about conditional probabilities has been shown. So, only marginal and joint probabilities can be used as known quantities. Every one of the subfamilies in the L_0-family

Table 2 Unique case for problems belonging to the $L_0C_0T_1$-subfamily

	A	$\overline{A}$	
B	$p(A\cap B)$		
$\overline{B}$	$p(A\cap \overline{B})$	$p(\overline{A}\cap \overline{B})$	
			1

indicates what information has to be given to solvers in the text of tasks and what has to be asked for (a conditional probability in all cases).

For example, the subfamily $L_0C_0T_1$ contains problems formulated with three joint probabilities. One conditional probability is asked for. If we want to formulate a task from problems in this subfamily, we have to select 3 joint probabilities from the 4 which are possible. Theoretically, there are 4 ways to do it or, which is the same, 4 cases of possible problems. However, for reasons of symmetry between the two basic events and the commutativity of the intersection between two events, it can be assured that there is only one basic problem in this family, for example, the problem formulated with these probabilities: $p(A\cap B)$, $p(A\cap \overline{B})$ and $p(\overline{A}\cap \overline{B})$, the other three cases being equivalent problems to this one. This case and the possible equivalence between problems can be better seen in a 2×2 table, as we show in Table 2.

For the unique case of the problem above, we have different options for asking questions about a conditional probability, so that together the case and the option give as a result a ternary problem of conditional probability. Due to the same reasons of symmetry as before and also for reasons of complementary probabilities, it is possible to reduce the options to two for this unique case of the problem. Thus, we may have only two basic problems in the $L_0C_0T_1$-subfamily and hence two basic tasks for every task variable which could be considered. These tasks could come from the case in Table 2 and from the question options $P_{\overline{A}}(B)$ or $P_B(A)$. For the first option, the complexity of the solution of the problem is 4 (three additive and one multiplicative relationships), while for the second one is 3 (2 and 1, respectively).

If we proceed following the method we have just shown above, the next table (Table 3) summarizes all the cases and possible options of ternary problems of conditional probability in the L_0-family.

Therefore, in this family of problems, one can consider a total of 6 basic cases of problems with 10 basic options for questioning. In fact, there are 10 basic problems if one basic case and a basic option are considered as a problem. If not then there are only 6 basic problems. With reference to tasks, there are 6 basic ones for each one of the task variables we are going to consider in their formulation.

There are also tasks in Table 3. They are formulated in a mathematic context. Quantities are expressed by means of the formal language of the theory of probability. So, a task like the following would be formulated in the most formal language possible in ternary problems of conditional probability:

Example 1 Given $p(A\cap B), p(A\cap \overline{B}), p(\overline{A}\cap \overline{B})$, calculate $P_{\overline{A}}(B)$ or $P_B(A)$ in which the task variables involved are the following: structural complexity is $L_0C_0T_1$; solution complexity is 4 or 3, depending on the question we ask; context is mathematical; data format for the quantities is probabilities.

Table 3 Problems of the L_0-level with their solution complexity degree

L_0-Family	Cases	Options	Solution complexity
$L_0C_0T_1$	$p(A\cap B)$, $p(A\cap\overline{B})$, $p(\overline{A}\cap\overline{B})$	$P_{\overline{A}}(B)$	4
		$P_B(A)$	3
	$p(A\cap B)$, $p(A\cap\overline{B})$, $p(\overline{A})$	Whatever it is	Indeterminate problem
$L_0C_1T_1$	$p(A\cap B)$, $p(A\cap\overline{B})$, $p(B)$	$P_{\overline{A}}(B)$	4
	$p(A\cap B)$, $p(\overline{A}\cap\overline{B})$, $p(B)$	$P_A(B)$	4
		$P_{\overline{A}}(B)$	3
$L_0C_2T_1$	$p(A\cap B)$, p(A), $p(B)$	$P_{\overline{A}}(B)$	3
	$p(A\cap\overline{B})$, p(A), $p(B)$	$P_{\overline{A}}(B)$	4
		$P_B(A)$	2
	$p(\overline{A}\cap\overline{B})$, p(A), $p(B)$	$P_A(B)$	4
		$P_B(A)$	4

However, in general, tasks are word-problems formulated in any context and using a data format for the quantities that is coherent with the chosen context. This choice implies that sentences referring to quantities in the text of problems also have to be coherent with the chosen context. Examples of problems from the L_0-family formulated in different contexts are the following:

Example 2 The fourth class of a secondary school has 30 pupils, boys and girls. Among these pupils, there are 7 girls who wear glasses, 10 girls who do not wear glasses and 8 boys who also do not wear glasses. Among the boys of the class, what percentage of them wear glasses?

Example 3 A group of people suffer from a skin infection. Some of them have been treated with an antibiotic and some not. 53 % of these persons have been cured. 35 % of these persons have been treated with the antibiotic and have been cured; 40 % of the persons have not been treated with the antibiotic and have not been cured. Among the persons who have been treated with the antibiotic, what is the percentage of those cured?

Example 4 A population at risk of suffering from AIDS undergoes a test to diagnose whether it has it or not. The test result is positive or negative in each case. The probability that one person at risk has AIDS is 0.57 and the probability that one person has a positive result is 0.47. One knows also that there is a probability of 0.23 that a person suffers from AIDS and gets a negative test result. For the persons who do not have AIDS, what is the probability that they have a positive test result?

The full description of these tasks can be seen in Table 4.

Table 4 Task description depending on task variables: structure, context, data format and complexities

Example	Structural complexity	Context	Data format	(Structural) solution complexity
1	$L_0C_0T_1$	Mathematic	Probabilities	3/4
2	$L_0C_0T_1$	Social-Stat.	Frequencies/Percentages for the question	4
3	$L_0C_1T_1$	Health-Stat.	Percentages	4
4	$L_0C_2T_1$	Diagnostic test in Health	Probabilities	4

The work which we have given examples for with problems from the L_0-family can be done with problems belonging to every family or subfamily. The main aim for doing this work is to have at our disposal a large set of problems, every one of them viewed as a canonical form of the class of problems that every family and subfamily represents. This set of problems could be used either for designing teaching models or for researching students' behaviour. Up to now, each of these possibilities has been tackled by teachers and researchers.

7 Research on Conditional Probability Problem Solving

Research into conditional probability problem solving is, therefore, only at the beginning, at least in the way we are proposing in this article. No previous research has been conducted in this field, despite Shaughnessy's (1992) recommendation which pointed out the reality of the difficulties of teaching probability by claiming, "teaching probability and statistics *is* teaching problem solving", thus encouraging researches to carry out research in this field. However, we did not have knowledge about the problems themselves, of how students solve these problems and of which factors may influence students' behaviour in solving them. With the acquisition of this knowledge researchers may help teachers in their task of improving students' competences in solving problems of conditional probability.

A few years ago, we started to conduct some research on conditional probability problem solving. Firstly, this was carried out in order to obtain knowledge about conditional probability problems, identifying a particular family of problems—ternary problems of conditional probability—as a model for many school tasks on conditional probability. Cerdán and Huerta (2007) modelled this family of problems by means of what they called trinomial graph of the *world* of ternary problems of conditional probability. This graph became a methodological tool for analysing and classifying problems (Lonjedo 2007), to study problems and contexts (Carles and Huerta 2007) and to represent students' resolutions of problems and determine students' competences and errors in the process of solving the problems (Edo et al. 2011). Secondly, the research was conducted in order to acquire knowledge of how students solved these problems, which were very hard tasks for all students at every school level and which students' difficulties in answering the problems correctly might be due to factors such as the task variables we mention in this work.

7.1 A Few Results from Studies About Problems' Difficulties

Based on previous research which found that students perform better if numerical data format of quantities is expressed in frequencies and percentages (Evans et al. 2000; Hoffrage et al. 2002; Girotto and González 2001; Zhu and Gigerenzer 2006; Huerta and Lonjedo 2006 and 2007), we designed research exploring difficulties of the L_0-Level ternary problems of conditional probability in a sample of 165 secondary school students (15–16 years old) without previous knowledge of conditional probability, and determined to what extent these difficulties are due to the task variables we considered in tests of problems (Carles et al. 2009). Each student solved 6 problems from this family, structurally isomorphic in pairs, with the context in which the problems were formulated being varied. Examples 2, 3 and 4 (Table 4) are examples of problems in the tests we used, although for the characteristics of the sample of students we decided that data format would be fixed for every quantity in the problems, frequencies for the known quantities and percentages for the unknown one which is always a conditional quantity. In this research, from students' solutions of the problems, the difficulties of the problems were defined and related to two task variables which were considered: context and data structure.

Because problem solving is a process (Fig. 1), the difficulties of the problems can arise anywhere during this process. In particular, we have found difficulties when the process starts with the reading and comprehension of the problem and students deciding to tackle the problem or not; difficulties after calculations, if any, when students were trying to give an answer to the question in the problem, whatever it was; and finally, when students gave the solution of the problem by means of a quantity. Therefore, we have a broader vision of the difficulties of a problem than is usually the case in research, which only pays attention to the number of the correct answers of an item or problem. Viewing the difficulties of the problems in the way which we propose, for a given problem, we can attach to it a new task variable that could be called the *difficulties of the problem* and investigate students' behaviour which might also be dependent on this new task variable.

After solvers read the problem they decided whether or not to tackle it. The number of students that did not tackle a problem gave us an initial idea of the difficulties of the problems. We define the *appreciated difficulty* (AD) of the problem as the percentage of students that do not tackle the resolution of a problem. Given a problem with an AD, whatever it is high or low, we always interpret the result as a consequence of the task variables, neglecting personal factors that we know can be influential on the students' decisions to start and continue the process of solving the problem or to abandon it, such us tiredness, having a negative attitude, emotions and feelings, and so on. Once the problem has been tackled, students end up giving an answer to the problem or not. This answer can be numerical or not, expressed by means of quantities or by subjective judgements. We call the percentage of students that tackled the problem but did not give an answer to the question the *problem difficulty* (PD). Once the answer to the question has been given, the solution must be a quantity. Of course, this quantity should have three components: number, description and format. Therefore, we measure the difficulty of giving the solution of a conditional probability problem by means of a quantity using two difficulties. The *solution difficulty* (SD) is the name we give to the percentage of students that answered the problem giving an incorrect number as the solution of the asked quantity. At the same time, in order to assess the difficulty of giving the second component of the asked quantity, for those students who gave an answer, we define the percentage of students who did not include a correct corresponding description of the given number in the solution as the *description difficulty* (DD). In particular, we define the percentage of students who gave a correct description (among those who gave any) as the *correct description difficulty* (CDD). Given a problem, these last difficulties measure to what extent it is difficult to answer the problem by not only giving a correct number in an appropriate data format but also its description. This is particularly interesting for those researchers keen in investigating how students express an asked conditional probability in a problem formulated in a particular context.

Carles et al. (2009) report that for problems from the L_0-family, formulated using frequencies for marginal and joint quantities and percentages for the asked conditional quantity, finding the correct solution of the problems is a very hard task for the 15–16 year old students (globally, AD = 28 %; PD = 18 %, SD = 69.9 %; DD = 45.3 % and DCD = 56.6 %). From the 990 analysed students' solutions, we can observe that students do not appreciate that tackling the resolution of these problems is a hard task, because a very high percentage of them tackle the problems and try to give an answer. However, what is really difficult for them is to give a correct answer to the asked conditional quantity by means of a correct numerical answer and its corresponding description. In fact, among all given answers to the problems, only 30 % of them were correct numbers (conditional percentage), described in 55.7 % of the cases by means of an expression. But among these, only 43.3 % had correct descriptions of the calculated conditional quantity.

For this sample of students and family of problems, difficulties varied depending on the contexts and data structure with which the problems were formulated. Thus, in this research the authors report that the context seems to be a more influential factor than the structure variable in some of the difficulties of the problems. While the structure variable seems to be a factor that has a meaningful influence on the appreciation and problem difficulty, though not so much on the solution difficulty of the problems, the context variable seems to be on all the calculated difficulties. In fact, this research reports that there are contexts that are significantly more influential on the difficulties than others. In particular, the context which they refer to as the most influential, *health-statistics* (see Example 3), could be included in a more general situation, on which Henry (2005) calls *Les situation causalités* in which there is a cause (to receive an antibiotic) that has an influence on the effect (to be cured of a disease). There is no sense in asking for a conditional quantity (probability) that people are not cured of a disease if they have an antibiotic, as many problems ask, leading to incorrect answers. The influence of the context on the difficulties is also observable in what is called *diagnostic test* context (see Example 4), where students can think that suffering from AIDS implies giving a positive result in a test, meaning that there is no sense in asking for a conditional probability relating these events. However, among the three contexts used in this research, the least influential is that which is called *social-statistics* (see Example 2) that could be included into the general situation that Henry (2005) calls *ensemble situation*, where the relationships between events is casual, as in the case of boys and girls who wear glasses or not.

Huerta and Cerdán (2010) and Huerta et al. (2011), in ongoing research, report that ternary problems of conditional probability are also very hard for prospective secondary school mathematics teachers. These students are college graduates with sufficient preparation in math to allow them to become mathematics teachers in secondary schools, after following year-long specific course in Mathematics Education.

54 students tried to solve a test consisting of 7 problems belonging to all four families in which we classify ternary problems of conditional probability, one problem from the L_0-family, two from L_1, three from L_2 and one from the L_3-family. All problems were formulated with quantities expressed in percentages according to the three contexts we used in previous research by Carles et al. (2009). All the problems were solved either by an arithmetical or probabilistic approach, depending on the students' decisions of how to tackle the resolution of the problems. Only one of them required algebra for its resolution, as all problems belonged to the L_3-family. Therefore, there were problems that are apparently more complex than others, both in their data structure and in their solution complexity.

In general, problems were not considered as difficult (on average, AP < 2 %) for this sample of students. But, in fact, the other difficulties were surprisingly high (on average PD 27.4 %, $\sigma = 14.8$, did not give an answer to the question in the problems) and it was even more surprising to discover that solution difficulties of the problems were so high (on average, SD over 69.6 %, $\sigma = 7.5$, and CDD $= 41$ %, $\sigma = 12$ %). This means that, due to the high dispersion in PD, there are problems which are more difficult to answer than others, but in almost all of them it is difficult for prospective mathematics teachers to find the correct answers to problems such as those we have analysed in this article. Furthermore, many of these answers were not quantities in the sense in which we have defined them, but were numerical answers without correct descriptions in 41 % of the cases.

7.2 Some Results About Students' Behaviour, Errors and Misunderstandings

Huerta and Lonjedo (2006, 2007) report that students' behaviour in solving conditional probability problems may depend on data format in expressing quantities in problems. Thus, they situate the students' reasoning in solving problems in a continuum from the arithmetical reasoning to the probabilistic reasoning depending on the role of the probabilities (as numerical data format) in the quantities used for solving the problem. If probabilities have no crucial role, if any, in the process of solving problems, then the reasoning is qualified as arithmetical, and if they do then the reasoning is qualified as probabilistic.

This distinction between students' arithmetical and probabilistic approach to solving conditional probability problems is also reported in Huerta and Cerdán

(2010). They observe that, for the prospective teachers who succeeded in solving the problems of the test, the most effective approach seemed to be the probabilistic one, although the most efficient was the arithmetical one using a 2×2 table. Because the problems were formulated in percentages, the probabilistic approach attempts to translate problems from the original formulation into problems in which quantities are expressed in the most formal probability language. The solving process requires using precise formulae to find the answer to the posed question. Many of the students who preferred the probabilistic approach did not succeed because of the formulae, either because they did not remember or misused them. Nevertheless, it seems that the most natural approach to all problems should be the arithmetical one.

In current research in which 15–16 year old students solve problems from the L_0-family, Edo designs a method to analyse students' resolutions of the problems based on trinomial graphs (Edo et al. 2011). For every problem in a test consisting of 6 problems from the L_0-family and from the students' written resolutions of every problem, Edo uses the graph in Fig. 8 as the space of problems, that is to say, the space where a student's resolution of a ternary problem of conditional probability may be produced. Therefore, she plots a graph representing every student's resolution of the problem. In order to do this, she makes analytical readings of the students' written resolutions, that is to say, paying attention to quantities and relationships between quantities used in the resolution of the problem. The resulting graph allows her to determine both competences in solving every problem, and hence strategies of resolution, and mistakes. She defines student's behaviour in a problem as competent if the quantities and relationships used in his/her resolution can be directly plotted onto the graph, independently of the number of the calculated quantities and relationships used in answering the question in the problem. Between competent behaviours in problem solving, she distinguishes the strategies of resolution by means of the number and type of intermediated quantities calculated and the relationships used to find those intermediated quantities and the solution to the problem.

Doing the translation from the written resolution to the trinomial graph, Edo identifies what she calls *quantity mistakes* and *relationship mistakes* in a student's resolution of a problem. A student makes a quantity mistake if any of the components of the quantity is wrong. She says that a student makes a relationship mistake between two quantities if in order to determine the new ones from them he/she uses a wrong relationship between them. In particular, she thinks that quantity mistakes can be produced both in the reading and comprehension phase and also in the calculation phase of new intermediate and final quantities. Thus, in the reading and comprehension phase, she says that if students misinterpret any of the known quantity in the text of the problem so that some of the components of the triple do not correspond to each other, then this is called an *interpretation mistake* of the quantity. She has identified four types of interpretation mistakes between quantities:

- Marginal as intersection;
- Intersection as conditional;

- Intersection as another different intersection;
- Intersection as marginal.

Besides this, she identifies the interpretation mistakes in the phase of calculation by means of the uses that students make of the known quantities for calculating the new ones. Thus, she observes misuses in some of the components of the calculated quantity as follows: the same number for two different descriptions; two different numbers for the same description; and finally, the discordance between the numerical component and the description and S, generally due to an expression mistake in describing the number of the quantity.

Due to the fact that in her research the question in all problems is about a quantity referring to a conditional percentage, she also identifies quantity mistakes in the solution phase of the problems. These mistakes are due to the fact that students generally give an answer to the question that is different from the conditional which is asked for. Thus, she identifies mistakes as follows:

- Giving as the answer to the problem the intersection or the marginal directly related to the conditional which is asked;
- Giving as the answer the inverse conditional; giving the marginal directly related to the inverse conditional.

Sometimes, students also give as the answer some of the calculated intermediate quantities not directly related to the asked conditional.

Finally, she says that a relationship mistake between two quantities has not taken place if the relationship produces a new quantity which makes sense in the process of solving the problem. If not, then a relationship mistake has been produced. Because all relationships in ternary problems of conditional probability are ternary relationships, every time students make a relationship between quantities that, in fact, are unrelated, there will be a relationship mistake. These relationships could be caused, basically, by the following: students relate two correct quantities giving an incorrect quantity (also making a quantity mistake); relate two quantities that are not related at all, a new but incorrect quantity will obviously be obtained.

8 Final Remarks and Further Questions

In this article, we propose a new perspective for research in probabilistic thinking from the point of view of problem solving. It is also an invitation to the community of mathematics educators and researchers to tackle probability problem solving researching and teaching, in a similar way as it has been done in arithmetical and algebraic problem solving.

Placing this research into the framework of the mathematical problem solving research requires taking into account the research variables (Kulm 1979) that should be considered in every piece of research into probability problem solving. Here, we have considered as research variables the following: content variable, structure variable, context variable and format variable, being conscious that other independent variables as, for example, semantic and syntactic variables, strongly related to language, are as important and also need to be considered. Only one of them has been fixed in this work: that referring to the content variable, the conditional probability. It is seen as a quantity in a problem either as a known or unknown quantity, or both. While solving problems, solvers are users of these quantities in their relationships with other quantities involved in the resolution of the problem. It is precisely the results of research into these uses, competent or not, of the conditional probability in a problem which could help us to understand the difficulties in teaching and learning the topic. Therefore, more research is needed.

Students' behaviour in solving a task seems not to be independent from the nature of the task they have to solve. So, for example, relating this behaviour to contexts, structures of data and language, in which tasks are put forward, either in an independent manner or considering everything together, also needs more research. What context would it be appropriate to consider? The decision with regards to this should take into account words said by Freudenthal with which we agree: "Probability applies in everyday situations, in games, in data processing, in insurance, in economics, in natural sciences. There is no part of mathematics that universally can be applied, except, of course, elementary arithmetic" (Freudenthal 1970), and the present day uses of the conditional probability.

What structure of data should be considered? For example, starting the research with problems such as we are suggesting in this work, ternary problems of conditional probability as precursors of other more complex problems, studying the students' behaviour depending on data structure of the problems?

What language aspects might be considered? For example, those aspects that could refer to the concept of conditional probability in various contexts and may cause students' interpretation mistakes and misuses, such as specific vocabulary in context: sensitivity, specificity, false positive, false negative, and so, in a diagnostic test situation; different grammatical forms to express a conditional quantity that may cause misinterpretation and confusion with joint (intersection) quantities in a given context?

What data format should be considered in the research? There is a possible answer: data format has to be coherent with the uses of the concepts in the context in which the task is formulated. Thus, in our view, solving tasks in varied data format (not necessarily given by probabilities) researchers might identify, describe and analyse the required critical thought referred by Freudenthal (1970) "In no mathematical domain is blind faith in techniques more often denounced than in probability; in no domain is critical thought more often required", as it is cited in Jones and Thornton (2005) who take Freudenthal's insight as a challenge and an opportunity for researchers (p. 84).

We would like to finish this work by referring to the teaching conditional probability problem solving. Again, we cite Freudenthal: "It should be made clear that the demand for technically formalized mathematics in probability is very low. Once you have mastered fractions, you can advance quite far in probability" (Freudenthal 1973, p. 583). So, why is solving ternary problems of conditional probability a very hard task for students, although in these problems the word probability is not explicitly mentioned? Is it caused by the low arithmetical and algebraic students' competence? Or, is it caused by the usual teaching models of probability, which are focused on techniques rather than on exploring context and phenomena referring to conditional probability through problem solving? In this work, we support this last position. Therefore, more research on problem solving based teaching models is also necessary, which might foster students' literacy in conditional probability and competency in solving problems in context. The theoretical model for the conditional probability will arrive later, if necessary.

References

Carles, M., & Huerta, M. P. (2007). Conditional probability problems and contexts. The diagnostic test context. In D. Pitta-Pantazi & G. Philippou (Eds.), *Fifth congress of the European Society for Research in Mathematics Education*, CERME 5 (pp. 702–710).

Carles, M., Cerdán, F., Huerta, M. P., Lonjedo, M. A., & Edo, P. (2009). Influencia de la estructura y el contexto en las dificultades de los problemas de probabilidad condicional de nivel N_0. Un estudio exploratorio con estudiantes sin enseñanza previa. In M. J. González, M. T. González, & J. Murillo (Eds.), *Investigación en Educación Matemática XIII* (pp. 173–185).

Cerdán, F. (2007). *Estudios sobre la familia de problemas aritmético-algebraicos*. Doctoral Dissertation, Universitat de València, Spain.

Cerdán, F., & Huerta, M. P. (2007). Problemas ternarios de probabilidad condicional y grafos trinomiales. *Educación Matemática*, *19*(1), 27–62.

Edo, P., Huerta, M. P., & Cerdán, F. (2011). Análisis de las resoluciones de problemas de probabilidad condicional mediante grafos. Un ejemplo. In M. Marín, G. Fernández, L. Blanco, & M. Paralea (Eds.), *Investigación en Educación Matemática XV* (pp. 337–350).

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

Fishbein, E. (1975). *The intuitive sources of probabilistic thinking in children*. Dordrecht: Reidel.

Freudenthal, H. (1970). The aims of teaching probability. In L. Rade (Ed.), *The teaching of probability and statistics* (pp. 151–168). Stockholm: Almqvist & Wiksell.

Freudenthal, H. (1973). *Mathematics as an educational task*. Dordrecht: Reidel.

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

Henry, M. (2005). Modélitation en probabilités conditionnelles. In M. Henry (Ed.), *Autour de la modelisation en probabilités* (Vol. 6, pp. 173–185). Besançon: IREM de Franche-Comté.

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

Huerta, M. P. (2009). On conditional probability problem solving research—structures and context. In M. Borovcnik & R. Kapadia (Eds.), Special issue on "Research and Developments in Probability Education". *International Electronic Journal of Mathematics Education*, *4*(3), 163–194.

Huerta, M. P., & Cerdán, F. (2010). El cálculo de probabilidades en la formación del profesorado de matemáticas de secundaria. In M. Moreno, A. Estrada, J. Carrillo, & T. Sierra (Eds.), *Investigación en Educación Matemática XIV* (pp. 353–364). Lleida: SEIEM.

Huerta, M. P., & Lonjedo, M. A. (2006). The nature of the quantities in a conditional probability problem. Its influence on the problem solving behaviour. In M. Bosch (Ed.), *European research in mathematics education IV. Proceedings of the fourth congress of the European Society for Research in Mathematics Education*, CERME (Vol. 4, pp. 528–538).

Huerta, M. P., & Lonjedo, M. A. (2007). The same problem in three presentation formats: different percentages of success and thinking processes. In D. Pitta-Pantazi & G. Philippou (Eds.), *Fifth*

congress of the European Society for Research in Mathematics Education, CERME 5 (pp. 732–741).
Huerta, M. P., Cerdán, F., Lonjedo, M. A., & Edo, P. (2011). Assessing difficulties of conditional probability problems. In M. Pytlak, T. Rowland, & E. Swoboda (Eds.), *Proceedings of the seventh congress of the European Society for Research in Mathematics Education*, University of Rzeszów, Poland (pp. 807–817).
Jones, G. A., & Thornton, C. A. (2005). An overview of research into the teaching and learning of probability. In G. Jones (Ed.), *Exploring probability in school (challenges for teaching and learning)* (pp. 65–92). New York: Springer.
Kulm, G. (1979). The classification of problem-solving research variables. In G. A. Golding & C. E. McClintock (Eds.), *Task variables in mathematical problem solving* (pp. 1–22). Columbus: ERIC.
Lonjedo, M. A. (2007). *Análisis de los problemas ternarios de probabilidad condicional de enunciado verbal y de sus procesos de resolución*. Doctoral Dissertation, Universitat de València, Spain.
Polya, G. (1957). *How to solve it*. Princeton: Princeton University Press.
Puig, L., & Cerdán, F. (1988). *Problemas aritméticos escolares*. Madrid: Síntesis.
Shaughnessy, J. M. (1992). Research in probability and statistics: reflections and directions. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 465–494). New York: MacMillan Co.
Tversky, A., & Kahneman, D. (1983). Evidential impact of base rates. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: heuristics and biases* (pp. 153–160). Cambridge: Cambridge Academic Press.
Zhu, L., & Gigerenzer, G. (2006). Children can solve Bayesian problems: the role of representation in mental computation. *Cognition*, *98*, 287–308.

Contextual Considerations in Probabilistic Situations: An Aid or a Hindrance?

Ami Mamolo and Rina Zazkis

Abstract We examine the responses of secondary school teachers to a probability task with an infinite sample space. Specifically, the participants were asked to comment on a potential disagreement between two students when evaluating the probability of picking a particular real number from a given interval of real numbers. Their responses were analyzed via the theoretical lens of reducing abstraction. The results show a strong dependence on a contextualized interpretation of the task, even when formal mathematical knowledge is evidenced in the responses.

Consider the conversation between two students presented in Fig. 1 and a teacher's potential responses. The scenario is a familiar one—two students grappling with opposing responses to a probability task. The task itself is less familiar—the likelihood of choosing a particular event from an infinite sample space.

In this chapter, we consider the specific mathematics embedded in the task in Fig. 1, and analyze the responses of practicing and prospective secondary school mathematics teachers as they addressed the scenario. Unlike conventional probability tasks, such as tossing a coin or throwing a die, a special feature of the presented task is that the embedded experiment—picking "any real number"—cannot be carried out. We begin by examining different aspects of probability tasks, the contexts in which they are presented, and the associated interpretations.

1 On Platonic vs. Contextualized

Chernoff (2011) distinguished between platonic and contextualized sequences in probability tasks related to relative likelihood of occurrences. He suggested that platonic sequences are characterized by their idealism. For example, when considering

A. Mamolo (✉)
University of Ontario Institute of Technology, 11 Simcoe St. N., Oshawa, ON, Canada, L1H 7L7
e-mail: ami.mamolo@uoit.ca

R. Zazkis
Simon Fraser University, 8888 University Drive, Burnaby, BC, Canada, V5A 1S6

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_34,

The following conversation occurred between Damon and Ava, two Grade 12 students. Imagine you are their teacher and that they have asked for your opinion. They approach you with the following:

Damon: I asked Ava to pick any real number between 1 and 10, write it down, and keep it a secret. Then we wanted to figure out what the probability was that I would guess right which number she picked.

Ava: Right. And we did this a few times. The first time I picked 5, and Damon guessed it right on the first try. The next time, since he said "any real number," I picked 4.7835. He never got that one.

Damon: So, we tried to figure out the probabilities. I think that the probability of picking 5 is larger than the probability of picking 4.7835. Ava thinks the probability is the same for both numbers. Who's right?

Please consider and respond to the following questions:

1. What is the probability that Damon would guess correctly the real number Ava picked between 1 and 10 when that number was:
   (a) 5? (b) 4.7835?
   How do you know?
2. Going back to Damon's question... Who is right? And also: Why are they right?

Fig. 1 The Task: Picking Any Real Number

a sequence of coin flips, it is assumed to be generated by an "ideal experiment—where an infinitely thin coin, which has the same probability of success as failure, is tossed repeatedly in perfect, independent, identical trials" (p. 4). In contrast, contextualized sequences are characterized by their pragmatism. For example, a sequence of six numbers when buying a lottery ticket is considered as contextualized.

In fact, most probability tasks that are found in textbooks or discussed in educational research pertaining to probability—such as tossing a coin, throwing a die, spinning a spinner—refer to contextual scenarios. However, it is a convention in mathematics, as well as in mathematics classrooms, to think of the events described in these tasks as platonic—as if they concern an infinitely thin coin, a perfect die, or a frictionless spinner. This convention is also accepted in educational research. Chernoff (2011) noted that early probability studies in mathematics education clarified the "platonic assumptions" in accord with mathematical convention, such as "fair coin, equal probability for Heads and Tails." However, such conventions have been taken for granted in subsequent research and specific assumptions are omitted.

Stepping aside from the conventional, Chernoff (2012) analyzed responses of prospective elementary school teachers to a contextual probability scenario: an answer key to a multiple choice test. His participants were presented with two different answer key options and were asked to determine which key was least likely. While the researcher interpreted the sequence platonically, considering both keys as "equally likely," the participants demonstrated a strong dependence on context in their responses. That is, their responses were contextualized: Regardless of their determination of which sequence was least likely (or if both were equally likely), the participants referred to their experiences with multiple choice tests either as a student taking a test, or as a teacher constructing the test. Their interpretation of the task "pragmatized" the sequence of outcomes rather than considered it platonically.

2 On Tasks, Experiments and Interpretations

While Chernoff (2011) labeled sequences as platonic or contextualized, in what follows we refine this distinction as it applies to probability tasks in general. Our refinement focuses on teasing out a distinction between how a probability task is presented versus how it is interpreted. For instance, the event of "tossing a coin and getting heads" is presented in a context, and as such the task—in which we are asked to determine the probability of this event—is contextual. However, the experiment itself and the resulting event can be seen as platonic under the assumptions listed above. This means that it is the *interpretation* of the experiment and of the event that is platonic, rather than the experiment itself. As such, we consider standard tasks used in probability classrooms and in probability research as "platonicized by convention." To reiterate, the "platonicity" of a coin-toss, or any other experiment, is a feature of an individual's interpretation of the experiment rather than a feature of experiment itself.

We also note that there are contextual events to which a platonic interpretation is not applicable. Consider, for example, tossing an uneven 11-sided solid with numbered sides and landing it on a 7, or meeting a high school friend in a foreign country. Since the solid is asymmetric, and the factors in "meeting a friend" are not predetermined, the probability of such events can be determined only experimentally or statistically.

An obvious question is whether there are probability tasks that are decontextualized, that is, not presented in some 'realistic' context. Indeed, a task can be presented in the following way: Consider events A and B, where $P(A)$ is $1/8$ and $P(B)$ is $1/3$. What is P (B and A)? While the probability here can be determined theoretically, this task can be contextualized by thinking, for example, of A as getting a 1 when rolling an octahedron with numbered faces and of B as picking a green marble from a bag that has 3 marbles, yellow, red, and green. However, such decontextualized tasks are out of scope of our current discussion.

We consider the task that is of interest in our study—that relates to picking a real number at random from a given interval—as platonic. As many other probability tasks, the scenario is described in a context, though unlike a coin toss, it cannot be carried out. As such, the experiment can only be imagined. The platonic interpretation of such an experiment, which is a mathematical convention, considers the infinite set of real numbers as a sample space, where each number has the same probability of being picked. As mentioned above, Chernoff (2012) has shown that when students are asked to consider tasks that fall outside of those platonicized by convention, their interpretation is contextualized, that is, embedded in their experience with the context of the scenario. We were interested in seeing whether individuals with strong mathematical preparation have a similar tendency towards contextualization, that is, whether their interpretation will be pragmatized when an "ideal experiment" is considered. Before exploring participants' interpretations, we take a closer look at the mathematics of our task.

3 Probability and an Infinite Sample Space

Recalling the scenario presented in Fig. 1, Damon and Ava argue about the likelihood of picking one real number versus another in the interval between 1 and 10. Mathematically, the probability of picking the number 5 is the same as the probability of picking the number 4.7835, even though pragmatically, the number 5 might be a more common choice. The conventional resolution hinges on the fact that each number occurs exactly once in the interval, and thus each has the same chance of being picked. But exactly what is this chance? The sample space in question is the set of real numbers between 1 and 10, written symbolically as $[1, 10]$. As there are infinitely many real numbers in that interval, the chance of picking 5, and the chance of picking 4.7835, is "1 out of infinity."

To be more specific, we first must take a brief diversion to the realm of infinity. By definition, the sample space $[1, 10]$ contains $\aleph_1$ many elements, where the symbol $\aleph_1$ (pronounced "aleph one") represents the transfinite cardinal number associated with any set of real numbers, including all of them. Transfinite cardinal numbers were defined by Cantor (1915) as analogues to the natural numbers (finite cardinal numbers). They describe the "sizes" of infinite sets, of which there are infinitely many. Cantor established definitions and algorithms for arithmetic with transfinite cardinal numbers, and as such, we may say that, by definition, 1 out of infinity, or more precisely, $1/\aleph_1$, is equal to zero. As such, the probability of picking 5, and the probability of picking 4.7835, is $1/\aleph_1 = 0$.

As with other aspects of transfinite cardinal numbers (see, e.g., Mamolo and Zazkis 2008; Tsamir and Tirosh 1999), dealing with probabilities and infinite sample spaces is paradoxical and counter-intuitive. As we allude to in our scenario, there is a paradox in the idea that the number 5 has a "zero chance" of being picked, and yet, it *was* picked. To give credence to this claim would take us beyond the scope of this chapter, but we mention it here as interesting mathematics trivia, and also as an illustration of why, in our view, probability questions that involve an infinite sample space must be interpreted "platonically"—not only is it impossible to carry out any such experiment, but the reality (if we can call it that) and the mathematics are not in accord with one another.

Paradoxical elements aside, there are other conceptual challenges associated with the statement "the probability of picking the number 5 is zero." Chavoshi Jolfaee and Zazkis (2011) observed that for a group of prospective secondary teachers the sample space of probability zero events was predominantly comprising "logically impossible" events, such as rolling a 7 with a regular die. They further noted confusion between an infinite sample space and a "very large" sample space, and as such between probability zero and probability that is "very small." Confusion regarding the distinction between a "very large, unknown number" and infinity (in the sense of transfinite cardinal numbers) is well documented (e.g., Mamolo 2009; Sierpinska 1987). This distinction coincides with what Dubinsky et al. (2005) considered process and object conceptions of infinity. Process and object conceptions of a mathematical entity are defined in the APOS Theory (Asiala et al. 1996) as corresponding to imagined actions and realized totalities, respectively. With respect to infinity, the process of counting numbers forever is juxtaposed with the totality of a set with infinitely many elements. In the context of our scenario, this distinction is significant as it impacts how the sample space, and thus the probability of the event, is treated—that is, whether the probability *approaches* zero, or *is* zero. The existence of a "realized totality" of infinitely many elements is strictly conceptual, and hence necessarily platonic. As such, to address probability questions with an infinite sample space, both the experiment and the event must be interpreted as platonic.

4 The Study

The task found in Fig. 1 was presented to six secondary school teachers, and was framed in the context of a teaching situation. For some, this framing elicited a 'teacher's response,' that is, rather than responding to the mathematics questions posed (what is the probability and why?), some participants responded with how they would address Damon and Ava had this scenario arisen within their classroom. These responses not only revealed participants' understanding and interpretation of the probabilistic situation, but also revealed their expectations regarding their own students' understanding of the associated ideas. The latter being outside the scope of our chapter, we focus our attention on participants' interpretations and resolutions of the probabilistic situation, highlighting prominent trends that emerged as participants attempted to make sense of the mathematics. These trends may be grouped under four main themes:

- Randomness and people's choices
- "Real" numbers
- Math vs. Reality
- The infinite and the impossible

As mentioned, the context of the task was pedagogical; it solicited responses to an imaginary conversation between two students. Furthermore, the mathematical problem that the students discussed we also interpret as contextual, where the familiar context is that of picking a number. However, the experiment itself we interpret as platonic since it can only be imagined and not actually carried out.

What we found in participants' responses to our task were different attempts to contextualize the problem, that is, to impose pragmatic considerations on both the experiment and the event. In particular, participants attempted to make sense of the task via contextualizing the experiment by considering, for example, what numbers individuals were likely to choose and why, or by considering what would happen if the experiment were actually carried out, and also via contextualizing the event by considering, for example, a context of infinity with which they were familiar. Table 1 below summarizes the trends in participants' responses within each of the aforementioned themes. In the following sections, we analyze these trends through the lens of contextualization.

Table 1 Themes and trends in participants' responses

	Ernest	Albert	Alice	Sylvia	Kurt	J.D.
Randomness and people's choices						
What A&D know about D&A	√	√	√			
What A&D would do		√	√	√		
People vs. Machines	√					
Likely numbers to choose	√	√	√	√	√	√
"Real" numbers						
For "average person"	√	√	√	√	√	√
In reality			√	√		
In classroom						√
$1/10$		√		√		√
Math vs. reality						
Experimental probability	√				√	
Theoretical probability					√	
"Should be" equal	√	√	√	√	√	√
Realistically $P(5) > P(4.7)$				√	√	
Experiment impossible				√	√	
The infinite and the impossible						
Almost impossible		√	√	√		
Approaches (or is?) zero	√	√	√	√		
$1/\infty$ or almost	√	√	√	√		
Limits			√			
Infinite sample space	√	√	√	√		

4.1 Randomness and People's Choices

The pragmatic considerations that participants imposed on the task in order to contextualize the experiment often hinged on the specific choices an individual was likely to make. Indeed, the majority of our participants attended to what numbers a Grade 12 student was likely to choose based on his or her experiences with numbers or, for example, based on how "well" Ava and Damon knew each other. For example, Ernest responded that "The answer to both parts (a) and (b) depends entirely on what Damon knows about Ava." He speculated that the two classmates "have some history together" and that Damon's guesses would depend on how familiar he was with his classmate; Ernest illustrated his position by indicating that if "Ava's favorite numbers are 5 and 7, and Damon is aware of this" then the probability will reflect this knowledge. Ernest argued that "we are dealing with people picking numbers between 1 and 10, and not machines" and that "most people think of numbers as discrete entities, and are terrified of the continuous values lurking between them." We return to the idea of how 'most people think of numbers' in a later section ("*Real*" *numbers*).

The notion that the probability would depend on whether Damon and Ava were humans or machines caught our interest, and when prompted to consider the task if the students were machines, Ernest reasoned that the probability would depend on what Damon was "programmed to pick." Ernest distinguished between a scenario where Damon "is programmed to pick only whole numbers, as was my [Ernest's] original interpretation" and a scenario where Damon "is programmed to pick any real number." In the latter case, Ernest acknowledged that

> If he is programmed to pick any real number, then there are an infinite number of choices, and it doesn't matter whether Ava picks a whole number or not. The probability they match is vanishingly small.

With the suggestion that the probability would be "vanishingly small" Ernest's response again foreshadows another trend in participants' reasoning as they dealt with an infinite sample space. We address this trend in the section *The infinite and the impossible*. In general, Ernest's resolution to the task hinged on who Ava and Damon were, interpreting the task in terms of the context of two classmates that know each other. This way of thinking about the task resonates with Nicolson's (2005) observation that consideration of how one throws a die (for example) influences children's interpretation of the probability of rolling a specific number. Sharma (2006) categorizes the intuition that how one handles a die impacts the randomness of a throw under the heading of "human control" and highlights the similar findings of several other researchers (e.g., Shaughnessy and Zawojewski 1999; Truran 1994). Who Ava and Damon are, and what they are likely to pick based on how well they know one another, is a contextualization of the experiment which draws on the idea of human control—Ava and Damon's choices are influenced (or controlled) by their "history together." Despite this contextualization, Ernest seemed to acknowledge that a platonic interpretation of the experiment was possible—if the two classmates were "machines". However, his scenario with machines was still contextualized, as the probability would depend on how the machines were programmed. Ernest also went on to wonder: "what if only one of them is a machine? Does it make a difference which one is the machine? Oh boy." For Ernest, the probability of guessing correctly depended on the individuals—whether they were machines or whether they were average Grade 12 students.

Considering what the 'average Grade 12 student' would guess or do influenced participants' interpretation of the experiment, as well as what the specific probabilities were and also how they compared. For instance, in determining the probability that Damon would correctly guess the real number Ava chose, Albert reasoned that:

> By asking for a number between 1 and 10, most people would instinctively choose a whole number, of which 5 would be a popular guess because it is in the middle. It is likely that Damon was aware of this. This results in a high probability that Damon would guess the number correctly, 1/10, or even better if you take into account that 5 is one of the more popular guesses.

Albert explained his response of 1/10 by arguing that "Most people would not relate to the term 'real' number" and that "had Ava been asked to pick a number between 1.000001 and 9.999987" the result would be different. Mathematically speaking, the fact that 5 is a popular number does not influence its probability of being chosen; however, Albert's interpretation of the experiment includes pragmatic assumptions of what Ava and Damon are likely to pick. We see similarities between Albert's response and responses characterized by the representativeness heuristic (Tversky and Kahneman 1974). As we elaborate upon in the following section, 'real' numbers for most people are, as Albert suggests, whole numbers. As such, 4.7835 may not be representative of a real number, and would thus have a less likely chance of being chosen. Albert explains that because whole numbers are more likely to be chosen, the probability would be "1/10, or even better"; however, if Ava and Damon are aware that they can pick non-whole numbers then a "different result" would have occurred. Specifically, Albert reasoned that once Ava realized she could pick "a number with decimals" then Ava "was surely not going to pick a whole number again (trying to make it more difficult for [Damon] to guess) and he knows it, so he would not guess a whole number either, and we are left with an infinite number of possibilities." As such, Albert concluded that "the probability of Damon guessing the number correctly became zero."

In Albert's contextualization of the experiment, he included considerations of Ava's intentions—that is, of picking a number that would be difficult for Damon to guess. This assumption may be a reflection of Albert's own goals in such an experiment, or in the 'average' individual's goals, but an intention to make the guess easy or difficult certainly was not implied in the phrasing of task. Of note is that Albert offered two different pragmatic interpretations of what the probability would be—one that restricted choices to whole numbers and popular guesses, and another that included consideration of the infinite sample space but that hinged on the intentions of the experimenter.

4.2 "Real" Numbers

As noted above, participants made reference to individuals' understanding of the term "real numbers." Ernest mentioned the "terror" of continuous values, Albert suggested that most people do not relate to the idea of "real numbers," and J.D. noted that even in an upper year high school mathematics classroom the term "real number" was not well understood. He responded:

> Judging that the students are in Grade 12, I would say the probability of Damon guessing five was 1 in 10. From my experience most 12th graders consider real numbers only as the set of integers, wholes and natural numbers.

Indeed, we see a conflict between the mathematical terminology and the 'everyday' terminology—in an 'everyday' context, there is nothing exactly "real" about an immeasurable number with infinitely-many non-repeating decimal digits. For most people, the numbers that are "real" are the natural ones. J.D. went on to reason that:

> As a Math teacher, I would say that Damon's probability of correctly guessing the real number Ava picked is 1 in 100,000. I correlated that with the number of place holders for remembering a number in short term memory (the most a normal person can remember without some sort of chunking strategy for remembering).

The relevance of students' understanding of real numbers and of how many digits a person can remember also influenced J.D.'s response to the second question in the task. He explained that:

> Ava is right that the probability is the same for both numbers. So long as the person answering the question has a strong understanding of what a real number means. This is a case where remembering a number in memory must have a finite answer. Short term memory stores can go anywhere from 5–9 digits, or possibly more with chunking strategies. Therefore, the probability of getting the right answer can range from 1 in 10,000 to 1 in 1,000,000,000.

J.D.'s contextualization of the experiment took into consideration that Grade 12 students were likely to guess whole numbers—as did Albert, and as such concluded that the probability is 1 in 10. However, J.D.'s interpretation of the experiment distinguished between his experience with what "most 12th graders consider real numbers" and with what he would say as a teacher of mathematics. Despite this distinction, J.D. nevertheless contextualized the problem by restricting the sample space to "the number of place holders for remembering a number in short term memory." For J.D. the possibility of choosing a particular number depended on the individual's ability to express that number in decimal digit form. In his perspective, "a number involves something with digits," and in his interpretation of the task, J.D. not only contextualized the experiment in terms of what a typical student might guess, but he also contextualized the experiment in terms of what he understands as numbers and how they are used in his classroom:

> Even when it is an irrational represented by a symbol such as π, the students in my classroom know for the most part what that symbol represents numerically. Calculators don't work on symbols. They need numbers at the very core. Thus, it is essential for the students to know numbers even irrationals (to a limit).

J.D.'s pragmatic interpretations included a sample space of only whole numbers (which are "real" for his students), as well as a contextual sample space of real-numbers-up-to-a-certain-digit that can be stored in memory. As JD thinks of numbers in terms of the digits in their decimal representation, it first appears from his response that he is confusing real numbers with rational numbers; however, later it becomes clear that he is thinking of rational approximations of real numbers.

4.3 Math vs. Reality

There were some participants who acknowledged that the task could have more than one resolution—a mathematical or theoretical one, and a 'realistic' one. For example, Alice reasoned that "the probability for guessing correctly is the same because each number is just one possible number that can be selected from all possible real numbers between 1 and 10." She wrote that the sample space has "infinitely many options. So I don't know how to express the probability." She guessed that the probability would "essentially tend towards 0," however:

> Realistically, I know it can't be 0 because that would be equivalent to saying it is impossible and we know it is definitely possible to guess the right number. So I would conclude that the probability of guessing the right number is finite.

Alice's considerations for what is realistic also emerged in her interpretation of the experiment. In resonance with Albert and J.D., Alice noted that:

> Realistically, however, I think someone is more likely to guess, or even select, a whole number when asked to pick any number. Why this is, I don't know. Maybe it's the way we think about numbers. Also, we use and are exposed to more whole numbers in reality. I also think that many people don't consider the mathematical definition of a real number when they refer to real numbers.

As before, what one is likely to guess or what one considers a real number is only relevant when assuming a pragmatic interpretation of the experiment—that is, when contextualizing the experiment. In a platonic interpretation, what an individual is likely to consider does not influence the probability of one number being chosen over another, as each number exists only once within the sample space. Alice emphasized that "people tend to gravitate towards whole numbers" because of what they "are exposed to more... in reality." As such, she concluded that "if we took a sample of, say, x students... we would likely find that the finite probability [from considering an infinite sample space] of guessing the right number, doesn't hold."

Kurt and Sylvia, who also distinguished between realistic and mathematical solutions, imposed pragmatic considerations on the experiment as if it could be carried out. Both also reasoned that the probability would depend on what type of numbers students were likely to choose; however, they also criticized possible implementations of the experiment. For instance, Kurt argued that the "experiment is flawed" and that "the experimental probability of the student guessing a natural number will be likely higher than choosing a real number with multiple digits behind the decimal." He reasoned that the experimental probability would not approach the theoretical one unless:

> the experiment is done fairly. In this case, fairly would require all the real numbers (from 1 to 10) to have an equal chance of being chosen (for example, writing each on a slip of paper and choosing one at random from a bag—this, of course, is not possible because we do not have a bag large enough to hold all slips of paper).

Similarly, Sylvia criticized the possibility of carrying out such an experiment noting that:

> for Damon to guess a real number that Ava has secretly chosen in her mind is a cruel task that may take him a lifetime. For example, Ava may choose the real number 4.82018392 and Damon is supposed to guess that within a reasonable amount of tries would almost be impossible. If Ava gave him hints as to whether his guesses were greater than or less than her number, Damon would stand a chance... Otherwise, I think, Damon is doomed.

Interestingly, while Kurt and Sylvia realized the physical impossibility of carrying out this experiment, they neglected to consider the situation platonically. Rather than re-interpret the experiment, they struggled to address issues of infinity from a pragmatic, 'realistic,' perspective, conceding impossibility and doom over an idealized interpretation of the task. This is consistent with naive resolutions to super-tasks such as the Ping-Pong Ball Conundrum (Mamolo and Zazkis 2008), where individuals are asked to imagine conducting an experiment with infinitely many ping-pong balls. Rather than grapple with a thought experiment, which implies a platonic interpretation of the task, individuals stressed the impossibility of placing so many balls in a large enough container. What is notable here is that our participants have substantial mathematics backgrounds and as such one might expect more readiness to interpret this situation platonically.

4.4 The Infinite and the Impossible

The previous themes considered participants' responses as they contextualized the experiment; however, we also note that participants who could provide a "mathematical" or "theoretical" response (i.e., who recognized a platonic interpretation of the experiment) were also likely to contextualize the event. That is, the event of picking the number 5 when picking any real number lead to a probability of zero—however, there is a conflict here, namely, how can the probability be zero when the event actually happened? This conflict was avoided by participants by contextualizing the imposing factor on the event—the infinite sample space. For example, Alice, who noted that "there are infinitely many options," tried to explain what it meant to determine the probability of an event given such a sample size. She wrote:

> Well, since probability can be expressed as the number of favorable outcomes over the number of possible outcomes, our probability will essentially tend towards 0, making it difficult to pick the right number (statistically speaking, anyways). This is because, 1 (favorable outcome) over a large number (a set of identified possible outcomes) will be a small number close to 0. If we identify more and more possible outcomes (which we know there will always be more), we are essentially replacing x with larger numbers, creating a question with limits.

Alice's notion of "infinitely many options" is in line with a process conception of infinity (Dubinsky et al. 2005). Alice identifies "more and more possible outcomes" rather than reasoning that infinitely many possible outcomes presently exist. To cope with this, she introduced a familiar for her context of infinity—that of limits. She explained further:

> We find the lim of $x \to$ infinity where $1/x$ represents the probability of picking the favorable outcome, tends towards 0 as the number of possible outcomes (x) increases. Realistically, I know it can't be 0 because that would be equivalent to saying it is impossible and we know it is definitely possible to guess the right number.

In contextualizing the event, Alice relied on a familiar mathematical context for the notion of infinity, that is, her experience with calculus and determining limits. This interpretation of infinity enabled Alice to sidestep the conflict of a probability zero event occurring by providing a context for which the probability could be interpreted as 'tending toward zero.' Such a notion resonates with Ernest's belief that "the probability [the numbers] match is vanishingly small" and with Albert's conclusion that the event "should be 'impossible' (probability zero or infinitely close to zero)." When pressed about his distinction between zero or infinitely close to zero, Albert emphasized that:

> The probability technically is not zero, it is infinitely close to zero! Zero would be if Ava was choosing a number and he was guessing colors or fruit. :-) So, it is not impossible for the event to occur... which is why I had "impossible" in quotes in my answer. The fact that he did it on the first try you can chalk up to a "fluke" factor.

This response resonates with Chavoshi Jolfaee and Zazkis's (2011) observation that probability zero events were predominately considered by prospective teachers as "logically impossible" events. (As previously mentioned, an event was considered as "logically impossible" when it was drawn outside of the sample space, such as getting a 7 when throwing a standard die.) Furthermore, the confusion that our participants experienced when trying to decide between "zero" and "close to zero" has its roots in a popular interpretation that conflates of the notions of a "probability zero event" with that of an "impossible event." However, mathematically, improbable does not imply impossible, but the subtle difference can be fully grasped only from the perspective of measure theory.

5 Discussion: On Contextualization and Abstraction

We investigated participants' responses to a probability task—determining the probability of a particular event drawn from an infinite sample space. Two particular events were presented:

- Picking 5 from the set of real numbers between 1 and 10, when picking a number at random
- Picking 4.7835 from the set of real numbers between 1 and 10, when picking a number at random

While mathematically both events have the same probability of occurring, and the participants acknowledged this fact, they also added explanations and constraints that suggested, either explicitly (like Sylvia and Alice) or implicitly (like Ernest and J.D.), that the first event was more likely than the second. This conclusion was rooted in participants' contextualized interpretation of the task. Some participants considered what numbers people might pick and how they interpret the game, others focused on a popular, rather than mathematical, interpretation of "real" numbers and on the impossibility of actually carrying out the experiment. Further, while participants commented that the probability of both events "should be" equal, the value assigned to this probability was "almost" zero, or "approaching zero," rather than "zero."

The framework of "reducing abstraction" introduced by Hazzan (1999) is applicable to account for the participants' responses. As individuals engage in novel problem solving situations, their attempts to make sense of unfamiliar and abstract concepts can be described through different means of reducing the level of abstraction of those concepts. Hazzan elaborated on three ways in which abstraction level can be interpreted:

(i) Abstraction level as the quality of the relationships between the object of thought and the thinking person.
(ii) Abstraction level as reflection of the process–object duality.
(iii) Abstraction level as the degree of complexity of the mathematical concept.

5.1 (i) Relationships Between the Object of Thought and the Thinking Person

Hazzan noted that the same object can be viewed as abstract by one person and concrete by another, that is, the level of abstraction depends on the person rather than the object itself. A powerful illustration of this idea is provided by Noss and Hoyles (1996) who suggested that "To a *topologist*, a four-dimensional *manifold* is as concrete as a *potato*" (p. 46). Hazzan and Zazkis (2005) further clarified "that some students' mental processes can be attributed to their tendency to make an unfamiliar idea more familiar or, in other words, to make the abstract more concrete" (p. 103).

Contextualization—which is embedding the experiment in a (familiar) context—can be seen as reducing the level of abstraction by moving from the unfamiliar (or less familiar and distant) to a familiar situation. Participants' responses that mention the impossibility of carrying out the experiment, that refer to a possible relationship between the two students Ava and Damon, that consider what people usually do when choosing numbers, or that consider real numbers by a rational approximation can be seen as illustrations of reducing abstraction in accord with interpretation (i). Further, we suggest that embedding the experiment in a familiar context can refer either to a "realistic" interpretation of the situation, or to a previous mathematical experience. The former is exemplified by responses such as Kurt's, wherein he suggested the experiment was flawed "because we do not have a bag large enough to hold all slips of paper (each with a real number written on it)." With respect to the latter we consider participants' responses that referred to familiar mathematical contexts in which they dealt with infinity, specifically the context of calculus and limits. Such a calculus-based contextualization (combined with a lack of exposure to measure theory or transfinite cardinalities) resulted in determining that the probability "approaches" zero, rather than is zero. This consideration manifests explicitly in Alice's response that "the probability of picking the favorable outcome, tends towards 0... [but] it can't be 0," and also underlies distinctions such as the one made by Albert that "the probability technically is not zero, it is infinitely close to zero."

5.2 (ii) Process–Object Duality

Researchers (e.g., Asiala et al. 1996; Sfard 1991) agree that process conception precedes object conception of mathematical notions and in such process conceptions can be viewed as less abstract. As mentioned above, process and object conceptions of infinity are juxtaposed as distinctions between how the sample space of real numbers may be interpreted. The interpretation of the infinite set of real numbers as a process, e.g., a set with indeterminate size and numbers that go on forever, or as an object, e.g., a completed set that contains infinitely many numbers, influences how the probability of the event of choosing a specific number from that set is described—either as *approaching* zero, or as *equal to* zero. Several responses of our participants (e.g., Alice) demonstrate a process conception of infinity and therefore are in accord with interpretation (ii). Further, an object conception of infinity goes hand in hand with a platonic interpretation of our task. As such any contextualization of the task which attempts to situate the experiment in terms of a process that could actually be carried out (e.g., such as Sylvia's) suggests an attempt to reduce the level of abstraction, and is also in accord with interpretation (ii).

5.3 (iii) Degree of Complexity of the Mathematical Concept

Infinity is a complex mathematical concept. Embedding infinity-related ideas in a probability situation adds further complexity. Hazzan (1999) relates the complexity of a mathematical entity to how compound it is, stating that "the more compound an entity is, the more abstract it is" (p. 82). As such, an individual may attempt to reduce the level of abstraction of a compound entity by examining only part of it. Hazzan exemplified that students employ this kind of reducing abstracting when thinking of a set in terms of one of its elements, as a set of elements is a more compound mathematical entity than any particular element in the set. In our case, participants demonstrated thinking of a sample space of real numbers by referring to subsets of the real numbers, such as natural numbers or numbers with a finite number of digits in their decimal representation. Thinking of a subset (and this particular case, a subset of lesser cardinality) is dealing with a less compound and more tangible object, and is in accord with interpretation (iii).

Hazzan and Zazkis (2005) note "that these interpretations of abstraction are neither mutually exclusive nor exhaustive" (p. 103). This observation is definitely applicable to our data. For example, referring to a familiar game of picking a number among natural numbers can be described in terms of interpretation (i) as well as interpretation (iii). Similarly, relying on calculus/limit interpretations of infinity corresponds to (i) as well as (ii). Hazzan developed the framework of reducing abstraction and showed its applicability to interpret undergraduate students' thinking when they struggle with difficult-for-them, at least initially, mathematical concepts. What is partially surprising is that in the cases described here, participants with rather strong mathematical backgrounds, who demonstrated their abilities to approach the task on a mathematical/theoretical level, also regressed to reducing abstraction and adding contextual considerations that were at times inconsistent with their formal mathematical solution.

6 Conclusion

Contextual considerations, are they an aid or a hindrance in mathematical problem solving? In mathematics, "or" is inclusive, and so is the case here. While contextualization may be helpful in the initial exposure to probability tasks, and it has been advocated that explicit experimentation aids students' learning (Biehler 1991; Stoll and Tarr 2002), there are also examples to the contrary.

In particular, our research considered responses of secondary school mathematics teachers to a contextual task, the probability of picking a given number from a specified interval of real numbers. However, explicit experimentation for this task is unattainable because the sample space under consideration is infinite; the experiment can only be imagined. We argued that when considering an infinite sample space a platonic interpretation of the experiment should be adopted. Yet, our data revealed that when dealing with an infinite sample space participants preferred a contextualized interpretation of the experiment. Namely, they included various pragmatic considerations in order to cope with the task, such as what would likely happen in reality (e.g., what numbers students would likely choose).

The tendency towards pragmatic considerations among participants with limited mathematics background was noted by Chernoff (2012); here we found a similar tendency among participants with rather strong mathematical preparation. Our data analysis demonstrated that contextualization hindered participants' ability to focus on the mathematics of the task, even when the needed knowledge was in their repertoire.

References

Asiala, M., Brown, A., DeVries, D., Dubinsky, E., Mathews, D., & Thomas, K. (1996). A framework for research and curriculum development in undergraduate mathematics education. In J. Kaput, A. H. Schoenfeld, & E. Dubinsky (Eds.), *Research in Collegiate Mathematics Education* (Vol. II, pp. 1–32). Providence: Am. Math. Soc.

Biehler, R. (1991). Computers in probability education. In R. Kapadia & M. Borovcnik (Eds.), *Chance encounters: probability in education* (pp. 159–211). Amsterdam: Kluwer Academic.

Cantor, G. (1915). *Contributions to the founding of the theory of transfinite numbers*. New York: Dover. (Trans: Jourdain, P., repr. 1955).

Chavoshi Jolfaee, S., & Zazkis, R. (2011). Probability zero events: improbable or impossible? In *Proceedings of 35th international conference for psychology of mathematics education* (Vol. 2, pp. 217–224). Ankara: PME.

Chernoff, E. (2011). Investigating relative likelihood comparisons of multinomial contextual sequences. In *Proceedings of CERME 7*. Available: http://www.cerme7.univ.rzeszow.pl/index.php?id=wg5.

Chernoff, E. (2012). Recognizing revisitation of the representativeness heuristic: an analysis of answer key attributes. ZDM. *The International Journal on Mathematics Education*, *44*(7), 941–952.

Dubinsky, E., Weller, K., McDonald, M. A., & Brown, A. (2005). Some historical issues and paradoxes regarding the concept of infinity: an APOS-based analysis: Part 1. *Educational Studies in Mathematics*, *58*, 335–359.

Hazzan, O. (1999). Reducing abstraction level when learning abstract algebra concepts. *Educational Studies in Mathematics*, *40*(1), 71–90.

Hazzan, O., & Zazkis, R. (2005). Reducing abstraction: the case of school mathematics. *Educational Studies in Mathematics*, *58*(1), 101–119.

Mamolo, A. (2009). Intuitions of 'infinite numbers': infinite magnitude vs. infinite representation. *The Montana Mathematics Enthusiast*, *6*(3), 305–330.

Mamolo, A., & Zazkis, R. (2008). Paradoxes as a window to infinity. *Research in Mathematics Education*, *10*(2), 167–182.

Nicolson, C. P. (2005). Is chance fair? One student's thoughts on probability. *Teaching Children Mathematics*, *12*(2), 83–89.

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

Sfard, A. (1991). On the dual nature of mathematical conceptions: reflections on processes and objects as different sides of the same coin. *Educational Studies in Mathematics*, *22*(X), 1–36.

Sharma, S. (2006). Personal experiences and beliefs in probabilistic reasoning: implications for research. *International Electronic Journal of Mathematics Education*, *1*(1), 33–54.

Shaughnessy, J. M., & Zawojewski, J. S. (1999). Secondary students' performance on data and chance in the 1996 NAEP. *The Mathematics Teacher*, *92*(8), 713–718.

Sierpinska, A. (1987). Humanities students and epistemological obstacles related to limits. *Educational Studies in Mathematics*, *18*, 371–397.

Stohl, H., & Tarr, J. E. (2002). Developing notions of inference using probability simulation tools. *The Journal of Mathematical Behavior*, *21*, 319–337.

Truran, K. M. (1994). Children's understandings of random generators. In C. Beesey & D. Rasmussen (Eds.), *Mathematics without limits. Proceedings of the 31st annual conference of the Mathematical Association of Victoria* (pp. 356–362). Melbourne: Mathematics Association of Victoria.

Tsamir, P., & Tirosh, D. (1999). Consistency and representations: the case of actual infinity. *Journal for Research in Mathematics Education*, *30*, 213–219.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, *185*, 1124–1131.

Cultural Influences in Probabilistic Thinking

Sashi Sharma

Abstract Decisions concerning business, industry, employment, sports, health, weather predictions, law, risk and opinion polling are made using an understanding of probabilistic reasoning. The importance of probability in everyday life and workplace has led to calls for an increased attention to probability in the mathematics curriculum. A number of research studies from different theoretical perspectives show that students tend to have conceptions about probability which impact on their learning.

The chapter has five sections. The first section outlines the importance of probability in both formal (school) and out-of-school situations and makes a case for teaching probability. The second section considers the different interpretations of probability. Although we use informal probabilistic notions daily in making decisions, research on probability has mostly focused on the classical and frequentist approaches, research on the subjective approach is almost non-existent. Further, the common culture may influence the informal ideas of probability. Yet, there appears to be minimal literature that deals with the educational implications of the role of culture. Hence, the third section draws on mathematics education research to discuss the interaction between mathematical cognition and social settings and culture. It will be argued that probability is no different and early notions as well misconceptions need to be addressed via this lens. It would help clarify the aims, purpose and limitations of probability education.

The next section will report on the effects of culture on students' probabilistic thinking. I will draw on examples from my work and few others who have studied cultural influences on probabilistic thinking to explain how probability is related to human culture and tied to cultural practices. The final section will consider the issues arising out of the literature and offer suggestions for meeting these challenges. Specifically, suggestions for teaching, assessment and further research will be outlined.

S. Sharma (✉)
The University of Waikato, Hamilton, New Zealand
e-mail: sashi@waikato.ac.nz

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_35,

1 Introduction

Again I saw that under the sun the race is not to the swift, nor the battle to the strong… but time and chance happen to them all. (Ecclesiastes 9:11, English Standard Version 2001)

King Solomon, the writer of the above proverb, perceived that chance (probability) is a force that none of us can avoid, it is an essential part of our everyday life. Probably the proverb dates from when the term chance was first being introduced. In the current world, probability not only has emerged as mainstream curriculum, it has become an essential component of numerate education (Gal 2005; Schield 2010; Zimmermann and Jones 2002). It is as important as mastering elementary arithmetic and literacy skills.

Probability influences how we make individual and collective decisions in everyday activities. Many games of chance are ruled by the roll of dice, toss of a coin or random numbers that computer games use. There are also the weekly lottos, "Big Wednesday Draws" and television games such as "Who wants to be a millionaire?" which has captured the attention of audiences all over the world (Quinn 2003). Many other areas of our lives are increasingly being described in probabilistic terms. We hear that a court case has been decided beyond "reasonable doubt." Doctors say that they are 95 % confident that a new medicine is effective. Decisions concerning risk (Gigerenzer and Edwards 2003; Kapadia 2009) and opinion polling are made using an understanding of probabilistic reasoning. As Zimmermann and Jones (2002, p. 222) write:

> *We are increasingly bombarded with vast numbers of facts and figures, it is imperative that we be able to engage in probabilistic reasoning to discriminate among substantive, misleading and invalid conclusions based on data.*

Gal (2002) writes that when students encounter probability messages in the media and other sources, they are required to interpret the language of chance, to be familiar with the notion of randomness and independence and to understand that events may vary and to critically evaluate the probabilistic statements they read. Moreover, Gal adds that anyone who lacks these skills is functionally illiterate as a productive worker, an informed consumer or a responsible citizen. Gal (2005) and Kapadia (2009) believe that it is essential to place sufficient emphasis on probability literacy in the real world.

From the above paragraphs, it can be argued that informal probability is firmly established in common culture. As such there is a potential for conflict and interactions between the knowledge and language of chance and probability which students acquire informally and mainly outside school and formal knowledge and rules that schools present. Formal probability is mostly introduced in the early school years as a model of chance in the context of random events which we hope students see as random. What if students do not see these events as random but on the contrary see outcomes of the roll of the die or toss of the coin as determined by God's will or by the person who throws it (Amir and Williams 1999; Sharma 1997). Then we may be building the model of probability on an insecure foundation.

Falk and Konold (1992) warn against an external view of probability. That is, learning of probability should not only pay attention to real-world demands. They propose that probability is a way of thinking and it should be learned for its own sake. Greer and Mukhopadhyay (2005) remind us that probability is a tool for understanding the physical and social worlds in contemporary society. It has been usually taught as part of the statistics strand and is often considered as the harder and less relevant topic. However, probability is an important learning area in its own right and usually contains the key concepts and processes of statistical thinking and learning more advanced subjects such as statistical significance or topics in other sciences.

Likewise, Franklin et al. (2007) state that probability is an essential tool in applied mathematics and mathematical modelling. It is an important part of mathematics education that enriches the discipline as a whole by its interactions with uses of mathematics. They add that probability helps us describe and quantify how things vary in real life situations (Franklin et al. 2005). So probability is not something that only helps us to gamble or play games of chance, it helps to investigate the real world we live in and respond critically to other peoples claims from investigations.

In recognition of the above considerations, there has been a movement in many countries to include probability at every level in the mathematics curricula.

In western countries such as Australia (Watson 2006), New Zealand (Ministry of Education 2007), and the United States (Franklin et al. 2007), these developments are reflected in official documents and in materials produced for teachers. In New Zealand, the mathematics learning area is now called "Mathematics and Statistics", and statistics is one of three content areas in the curriculum document. Probability underpins the other two strands of statistics (statistical thinking and statistical literacy) in the New Zealand curriculum (Ministry of Education 2007). In line with these moves, Fiji has also produced a new mathematics prescription at the primary level that gives more emphasis to statistics at this level (Fijian Ministry of Education 1994). However, probability is first introduced in the early years at the secondary level. The use of relevant contexts and drawing on students' experiences and understandings is recommended for enhancing the students' understanding of probability (Ministry of Education 2007; Watson 2006). Clearly, the emphasis is on producing intelligent citizens who can reason probabilistically and evaluate statistical information that permeate their lives and environment.

This chapter has five sections. The first section outlines the importance of probability in both out-of-school and formal (school) situations and makes a case for teaching probability. The next section considers the issues in probability education. Specifically, different interpretations of probability are considered: theoretical, frequentist and subjective. Although we use informal probabilistic notions daily in making decisions, research on probability has mostly focused on the classical and frequentist approaches, research on the subjective approach is scare. Further, the common culture may influence the informal ideas of probability. Yet, there appears to be minimal literature that deals with the educational implications of the role of culture in probability education.

The third section draws on mathematics education research to discuss the interaction between mathematical cognition and culture. It will be argued that probability is no different, and early notions as well as misconceptions need to be addressed via the cultural lens. This would help clarify the aims, purpose and limitations of probability education.

The next section will report on research on the effects of culture on students' probabilistic thinking. I will draw on examples from my work and few others who have studied cultural influences on probabilistic thinking to explain how probability is related to human culture and tied to cultural practices. In particular, cultural beliefs (religion, superstitions, language and experiences) strongly affected student thinking.

The final section will consider the issues arising out of the literature and offer suggestions for meeting these challenges. Specifically, suggestions for teaching and further research will be outlined.

2 Issues in Probability Education

Research shows that many students find probability difficult to learn and understand in both formal and everyday contexts, and that we need to better understand how learning and understanding may be influenced by ideas and intuitions developed in early years (Barnes 1998; Chiesi and Primi 2009; Fischbein and Schnarch 1997; Jones et al. 2007; Sharma 1997; Shaughnessy and Zawojewski 1999). Shaughnessy and Zawojewski (1999) report that even when secondary students were successful on such items as identifying the probability of a simple event on the National Assessment of Educational Progress (NAEP) study, they experienced difficulty applying this information in problem-solving situations. Additionally, there are different viewpoints on how to teach probability best so that students leaving school may be able to interpret probabilities in a wide range of situations (Jones et al. 2007; Kapadia 2009; Stohl 2005). These views are based on different interpretations of probability.

2.1 Interpretations of Probability

People think about probability in at least three different ways (classical, frequentist and subjective) and these views can be manifested in the teaching and learning process. Each of these interpretations has its advantages and disadvantages. If students are to develop meaningful understanding of probability, it is important to acknowledge these different interpretations and to explore the connections between them and the different contexts in which one or the other may be useful.

The classical (theoretical) viewpoint assumes that it is possible to represent the sample space (all possible outcomes) as a collection of outcomes with known probabilities (Sharma 2009). When the outcomes are equally likely, the probability can be found by counting the number of favourable outcomes and dividing by the total number of outcomes in the sample space. For example, the probability of rolling a "six" on a regular six-sided die is one-sixth. Additionally, one can examine the symmetry of a regular six-sided die and estimate the probability of rolling a "six" as one-sixth. In both cases, the theoretically derived probability is an estimate of the actual probability that is not known (Stohl 2005). Batanero et al. (2005) argue that although equiprobability may be clear when rolling a die or playing a chance game, it is not the same in complex everyday situations. For example, equal chance can hardly be found in rare cases such as weather predictions, risks and epidemics. Furthermore, in some classroom situations, there is no theoretical probability. Indeed, when rolling an unfair die, the only way of estimating the actual probability of an event is to perform an experiment using a large number of trials. Stohl (2005) writes that the inability to predict whether or not a four will occur on a roll of a die may be due to the inability of people to directly account for complex nature of physics involved. Indeed, we cannot exactly determine the actually probability of rolling a four. Dollard (2011) writes that theoretical probability is commonly used in classrooms because it is easily applied to random chance devices such as dice and spinners where the definition of the sample space in terms of equally likely outcomes is relatively straightforward and it allows teaching to avoid the uncertainty of real random events.

The experimental or frequency interpretation assumes that the probability of something happening can be determined by doing experiments. A large number of identical trials (e.g. tossing 2 coins) are conducted, and the number of times a particular event (e.g. 1H and 1T) occurs is counted. The greater the number of trials, the closer the experimental probability will move towards the theoretical probability of an event. Moreover, the frequentist approach defines *probability as the hypothetical number towards which the relative frequency tends when stabilizing* (Batanero et al. 2005, p. 23). By comparing inferences from their theoretical and empirical work students can evaluate and modify their hypotheses. From a practical interpretation, the frequentist approach does not provide the probability of an event when it is physically impossible to repeat an experiment a very large number of times. For example, there is no possibility of conducting repeated trials for estimating the probability that one will live beyond 70 years or that one's house will be burgled within a year. It is also difficult to decide how many trials are needed to get a good estimation for the probability of an event. We cannot give a frequentist interpretation the probability of an event which only occurs one time under the same conditions. According to Batanero et al. (2005, p. 23), *the most significant criticism of the frequentist definition of probability is the difficulty of confusing an abstract mathematical object with the empirical observed frequencies, which are experimentally obtained*. In addition, this raises the problem of confusing model and reality and makes the modelling process difficult to understand for students who need to use abstract knowledge for probability to solve concrete problems. As Greer and Mukhopadhyay (2005, p. 316) pointed out, *a major error in teaching probability is to pretend that all situations modelled probabilistically are clean*.

The subjective view regards probability *as a numerical measure of a person's belief that an event will occur based on personal judgement and information about an outcome* (Jones et al. 2007, p. 913). Hence, it may not be the same for all people. This approach depends on two aspects, the event whose uncertainty is contemplated and the knowledge of the person estimating the probability. There are two aspects of subjective notions: one relates to intuitive conceptions, the other is that many applications involving risks probability are neither objective nor open to frequency interpretation. Although we use informal probabilistic notions daily in making decisions, curriculum documents are largely silent on the kind of knowledge base students are expected to use in making decisions (Kapadia 2009). Many of the ideas that students bring to the classroom belong to this third category. This section indicates the multifaceted nature of probability; it also suggests that *teaching cannot be limited to one of these different perspectives because they are dialectically and experientially intertwined* (Batanero et al. 2005, p. 31). Research on probability measurement has mostly focused on the classical and frequentist approaches. Falk and Konold (1992) could not locate cognitive research on the subjective approach to probability measurement.

In mathematics education research, it is recognised that early conceptions of ideas need to be discussed so that children can develop better understanding (Konold 1991; von Glasersfeld 1993). Probability is no different, and early notions as well misconceptions such as representativeness (Kahneman et al. 1982) need to be addressed during instruction (Barnes 1998; Kapadia 2009; Sharma 2009). According to Kapadia, taking a subjective approach would help clarify the aims, purpose and limitations of probability calculations and models. Kapadia adds that *probability programmes are usually based on relatively formal mathematical notions of probability, where contexts and underlying subjective knowledge is to be ignored* (p. 375).This may in part be due to the fact that it is easier to set questions in theoretical and frequentist contexts. Although the questions may be reliable and consistent in testing probability ideas accurately students may not be able to apply these ideas in real life situations. Explicit reference needs to be given to subjective notions in curriculum documents.

2.2 Lack of Cultural Research

In some countries like New Zealand (Averill et al. 2009) and USA (Lubienski 2002; Nasir et al. 2008), pressure has mounted to reflect in the school curriculum the multicultural nature of their societies and there has been widespread recognition of the need to re-evaluate the total school experience in the face of the education failure of many children from ethnic minority communities (Strutchens and Silver 2000). In other countries, like PNG and Fiji, there is criticism of the colonial or western educational experience and a desire to create instead an education which is relevant to the home culture of society. In all of these cases, a culture-conflict (Bishop 1994; Lubienski 2002) situation is recognised and curricula are being re-examined. Tate (1997) and others have argued that ignorance of the cultural diversity of students has contributed to the underachievement of some minority groups in mathematics.

Shaughnessy (1992) raised concerns about the lack of research in probability education outside of western countries. He advocated large and small scale studies that examined group and cultural differences on students' thinking in decision making and probability estimation tasks. He indicated that it would be interesting to determine how culture influences conceptions of probability, whether biases and misconceptions of discussed in literature are artefacts of western culture or whether they vary across cultures. After almost a decade, Greer and Mukhopadhyay (2005) argue that although probability is a cultural construction, historically linked to religious and philosophical issues, it has been introduced with minimal regard to cultural background of students. They conjecture that an almost homogeneous probability curriculum that exists in many countries (mostly imported from western countries) does not address whether the culture of the participants might have any effect on their probabilistic thinking. Additionally, the separation of school as an activity system from out-of-school systems plus the type of items used for assessment combine to mask any effects of the students' background in such studies. Greer and Mukhopadhyay (2005) claim that *sensitivity to the culture of others is needed when teaching mathematics in multicultural classes and when teaching a curriculum that is not indigenously grounded* (p. 318). Some research (Chiesi and Primi 2009; Peard 1995; Polaki 2002) has been undertaken outside of western countries. Chiesi and Primi (2009) presented views of development of probabilistic reasoning that runs contrary to what is described by developmental theories (Piaget and Inhelder 1975). They claim that aspects of college students' performance seem to get worse with age. Polaki (2002) conducted teaching experiments on the probabilistic reasoning of elementary students in Lesotho. Peard (1995) investigated the effect of social background on high school students' probabilistic thinking. He found significant probabilistic intuitions for participants whose social background included extensive familiarity with track racing. According to Greer and Mukhopadhyay (2005), these studies relate strongly to American or English research and do not focus on how the culture of the subjects may have impacted on their decision making. Greer (2001) points out that Fischbein did refer to the experiences related to the geographical and cultural environment of the individual and to the particular practice of domains of the individual's life; however, he did not include these aspects in his experiments.

3 Culture and Mathematics Education

The emergence of probability from games of chance (Batanero et al. 2005; Bishop 1988a, 1988b; Gabriel 1996; Greer and Mukhopadhyay 2005) and *divinatory predictions that have arisen from magical ancestral way of thinking* (Batanero et al. 2005, p. 15) make it clear how probabilistic thinking is culturally embedded. For example, Gabriel discussed the complex relationship between gambling and spirituality among indigenous American people. Indeed, this cultural knowledge is important when teaching students from this community. However, there appears to be minimal literature that deals with the educational implications of the cultural perspective. This section will draw on relevant literature from mathematics education research to explain these issues.

There are several obstacles facing any researcher who wishes to examine issues in statistics education through a cultural lens. Culture is situated at the intersection of several areas of inquiry, including those involving race, power and social class. This complexity poses two major challenges.

First, because of the specialised nature of academic fields, statistics education researchers tend to be unfamiliar with current research and theories relating to culture and power. Indeed, a researcher wishing to conduct cultural studies of statistics classrooms needs to be grounded not only in statistics education research but also in research from fields such as linguistics, anthropology and sociology. This is particularly difficult because the perspectives guiding studies of culture in these fields have been both highly contested and shifting (Bishop 1988a, 1988b; Nasir et al. 2008; Vithal and Skovsmose 1997; Wax 1993; Zaslavsky 1998). Indeed, in a reaction against cultural-deficit theory, scholars have promoted cultural difference model in which the language and practices of underserved students need to be considered in the classroom.

Secondly, mathematics and statistics curricula have been slow to change due to a primarily popular and widespread belief that mathematics is culture-free knowledge (Benn and Burton 1996; Bishop 1988a, 1988b). Indeed, the argument is that a negative times a negative gives a positive wherever you are, and triangles all over the world have angles which add up to 180 degrees (Bishop 1988a, 1988b). However, as Bishop points out, as soon as one begins to focus on the particulars of these statements, one's beliefs in universality of mathematics tends to feel challenged.

In the last decade, the field of research into the role of culture in mathematics education has evolved from ethnomathematics, critical mathematics education, and everyday mathematics, although these perspectives are connected (Bishop 1988a, 1988b; Moreira 2002; Nasir et al. 2008; Vithal and Skovsmose 1997). Ethnomathematics is the study of cultural aspects of mathematics (D'Ambrosio 2001). It encompasses a broad cluster of identifiable groups, ideas ranging from distinct numerical and mathematical systems to multicultural mathematics education (Zaslavsky 1998). From these perspectives, an attempt has been made to develop an alternative mathematics education or culturally responsive pedagogy which expresses social awareness and political responsibility (Averill et al. 2009; Martin 2006).

The notion of culture has been contested and given varied and sometimes competing interpretations in the literature (Nasir et al. 2008; Vithal and Skovsmose 1997; Zaslavsky 1998). Let us start by defining what we mean by culture. Bishop (1988a, p. 5) favoured the following definition of culture:

> *Culture consists of a complexity of shared understandings which serves as a medium through which individual human minds interact in communication with one another.*

Presmeg (2007) stated that while the above definition highlights the communicative function of culture that is particularly relevant in teaching and learning mathematics, it does not focus on the dynamic view of culture that results in cultural change over time due to movement between different social groups.

Presmeg viewed culture as *an all-encompassing umbrella construct that enters into all the activities of humans in their communicative and social enterprise* (2007,

p. 437). Presmeg adds that researchers may speak of the culture of a society, of a school, and culture in all these levels of scale can impact on the mathematical learning of students. Presmeg points out that the mathematics classroom itself is one place in which culture is contested, negotiated and manifested.

I draw on a perspective of culture with roots in Vygotskian theory of learning (Vygotsky 1978). In this perspective, culture can be described as a group of people's shared way of living. It is used to encompass commonly experienced aspects of the groups' lives such as shared knowledge, backgrounds, values, beliefs, forms of expression and behaviours that impact classroom interactions (Nickson 1992). This means that the cultural practices that we engage in as we move across everyday, school and professional contexts both shape and constitute our learning. However, it can become a complicated concept in schools in which school and classroom cultures exist within broader cultural contexts. Students' home and school cultures may be very different from one another which means students need to operate differently in these contexts while for other students these cultures may be more compatible (Abbas 2002; Averill et al. 2009; Bourdieu 1984; Clark 2001).

The above definition resonates with principles of socio-cultural theories combined with elements of constructivist theory which provide a useful model of how students learn mathematics. Constructivist theory in its various forms is based on a generally agreed principle that learners actively construct ways of knowing as they strive to reconcile present experiences with already existing knowledge (Konold 1991; von Glasersfeld 1993). Students are no longer viewed as passive absorbers of mathematical knowledge conveyed by adults; rather they are considered to construct their own meanings actively by reformulating the new information or restructuring their prior knowledge through reflection (Cobb 1994). This active construction process may result in misconceptions and alternative views as well as the students learning the concepts intended by the teacher.

Another notion of socio-cultural theory derives its origins from the work of socio-cultural theorists such as Vygotsky (1978) and (Lave and Wenger 1991) who suggest that learning should be thought of more as the product of a social process and less as an individual activity. Learning is viewed as developing greater participation in both in-school and out-of-school contexts. There is strong emphasis on social interactions, language, experience, collaborative learning environments, catering for cultural diversity, and contexts for learning in the learning process rather than cognitive ability only. Students learn through social interaction, by talking, explaining, listening and actively exploring concepts with their peers in whole class and small group situations.

The dilemma is how to incorporate cultural (out-of-school) mathematics in school mathematics classrooms in ways that are meaningful to students and that do not trivialise the mathematical ideas inherent in those practices. This issue remains a significant one for mathematics education research. Nasir et al. (2008), as part of a study on thinking and learning across contexts, asked students to solve average problems in two ways. In one set of tasks, the problems were framed by basketball practices, and others were given in the format of a typical school mathematics worksheet. The researchers report that some of the subjects demonstrated rich mathematical problem solving strategies in non-school context in a form markedly different from what we typically consider school knowledge. These experiences led these researchers to reject the notion that knowledge is independent and hence transportable. Similar findings have been reported Lave et al. (1984) and De Abreu et al. (1992). Lave et al. (1984) explored everyday mathematics, examining the strategies people use when they go grocery shopping. They reported differences between the nature of students' solutions to routine problems and strategies taught in schools.

De Abreu et al. (1992) documented that students who participated in mathematics practices of their communities and who also attended school did not view both of these practices as having the same status. They interviewed Severina, a 14 year old daughter of a sugar cane worker in Brazil. Severina experienced a strong discrepancy between her everyday life and school mathematics. Although she thought her calculations were done in a proper way (it reflected her teacher's way), she considered her fathers strategy would give a correct answer because it corresponded to their everyday experiences.

McKnight et al. (1990) recognised the difficulties involved in connecting 'everyday' non-mathematical or non-statistical components in the interpretive media graphs. They wrote:

> *Translation from the 'clean' world of abstract mathematics to the 'messy' world of everyday reality—in which all of our knowledge has links to other knowledge as well as links to personal beliefs and emotional reactions—introduces yet another complexity. Sometimes that other knowledge—or what one thinks is other relevant, linked knowledge—or those beliefs and affective reactions interrupt the more cognitive, information processing tasks of interpreting the graph* (p. 14).

From the cultural view of mathematics, there are requirements of what it means to know mathematics. In addition, there are entailments of both what it means to be a mathematics learner and what it means to be an effective mathematics student in a particular community. They need to manage the structures and discourses of everyday or cultural practices versus classroom mathematical knowing or school mathematics (Abbas 2002; Averill et al. 2009). Based on these studies, some argue that teaching should seek to better contextualise and make mathematics relevant to students' real life. The body of research on the relationship between cultural knowledge and domain knowledge provides an important lens through which we can understand and study school statistics classrooms which are sometimes considered culture-free.

4 Research in Probabilistic Thinking

It is often taken for granted that children see common devices such as dice, coins and spinners as random. However, research shows that a number of students think that their results depend on how one throws or handle these different devices (Amir and Williams 1999; Fischbein et al. 1991; Truran 1994; Shaughnessy and Zawojewski 1999). In other words, there is a belief that the outcomes can be controlled by the individuals or some outside force. Fischbein et al. asked 139 junior high school students (prior to instruction) to compare the probability of obtaining three fives by rolling one die three times versus rolling three dice simultaneously. Two main types of unequal probabilities were mentioned by about 40 % of the students. Of these students, about three-fifths considered that, by successively throwing the die, they had a higher chance of obtaining the expected result, and about two-fifths considered that by throwing three dice simultaneously they would have a higher chance of obtaining the expected results. Some of the justifications provided by the students for successive trials were:

> By rolling one die at a time, one may use the same type of rolling because the coins do not knock against each other and therefore do not follow diverse paths.

The opposite solution, that is, there was a greater chance of getting three heads by tossing the coins simultaneously also had its proponents. Reasons given included:

> Because the same force is imparted.
> One can launch in the same way.

The explanations indicate a belief that the outcomes can be controlled by the individual.

In the Truran studies (1994), many different methods of tossing coins and dice were described by children in order to get the result they wanted. For example, for tossing three dice together, some children thought that it is better to throw the dice one at a time because (when tossed together) dice can bump into each other and change the numbers which would otherwise have come up. Even if their carefully explained and demonstrated method did not work, children were still convinced that if they did everything right, it would work the next time. With respect to the task:

> Are you more likely to get 5 on each of these three dice, by rolling one dice three times or rolling all three dice together? Can you say why?

the responses suggest that the children's perceptions of the behaviour of random generators with similar structure were based on the physical attributes of the random generators. As one child explained (Truran 1994),

> Tossing three dice together, the dice rub against each other and that makes the numbers change; it is better to throw them one at a time.

Shaughnessy and Zawojewski (1999) documented similar statements among 12th graders on probability tasks. Rather than using the sample space or multiplication principle, the students attributed to some sort of physical property of the spinners such as the rate at which the spinners were spun, the initial position of the arrows, or even some external influence, for instance, the wind.

Zimmermann and Jones (2002) studied high school students' thinking and beliefs by confronting them with problems involving two-dimensional probability simulations. Various student beliefs emerged from their analysis of the data. They categorised these into helpful or problematic beliefs. What they found disturbing was the fact that some students believed, albeit to different degrees, that simulation cannot be used to model a real-world probability problem. The results of this study indicate that the feasibility of modelling a real life situation needs to be discussed carefully with students.

Amir and Williams (1999) propose that beliefs appear to be the elements of culture with the most influence on probabilistic thinking. They interviewed 38 11- to 12-year-old children about their concepts of chance and luck, their beliefs and attributions, their relevant experiences and their probabilistic thinking. Some pupils thought God controls everything that happens in the world while others thought God chooses to control, or does not control anything in the world. Several pupils believed in superstitions, such as walking under a ladder, breaking a mirror and lucky and unlucky numbers. There were also beliefs directly related to coins and dice; for instance, when throwing a coin, tails is luckier. A majority of children in the Amir and William study concluded that it is harder to get a 6 than other numbers (17 out of 21 interviewees). The children remembered from their experience of beginning board games waiting a long time for a 6 on the die. They conjectured that while many of the children attributed outcomes of chance to God, *two separate modes of thinking and reasoning are simultaneously present without damaging interactions of interference* (p. 95).

Sharma (1997) explored Fijian–Indian high school students' ideas about probability and statistics in a range of constructs. In this section, the focus is on equally likely and independence constructs. The sample consisted of a class of 29 students aged 14 to 16 years of which 19 were girls and 10 were boys. According to the teacher, none of the students in the sample had received any in-depth instruction on probability. In Fiji, probability is introduced from the theoretical perspective.

In situations such as tossing a coin (a one-stage experiment), the outcomes are said to be equally likely, the particular one that does occur when a coin is tossed being purely a matter of chance. Thus, a head is said to have a 50–50 chance, or a 50 % chance, or one chance in two of occurring. The single die (Item 1A), the advertisement regarding the sex of a baby (Item 1B) and the Fiji Sixes task (Item 1C) were used to explore students' understanding of the equally likely concept.

Item 1A: Single die problem

Manoj feels that a six is harder to throw than any other number on a dice?
What do you think about this belief? Explain your thinking.
Item 1B: Advertisement involving sex of a baby

Expecting a baby? Wondering whether to buy pink or blue?
I can GUARANTEE to predict the sex of your baby correctly.
Just send $20 and a sample of your recent handwriting.
Money-back guarantee if wrong!
Write to. . . . .

What is your opinion about this advertisement?

Item 1C: Fiji Sixes task (Fiji sixes is a lotto game)
Here is a Fiji Sixes card filled out by a friend of mine. She had to select six numbers out of 45 so she chose the six consecutive numbers starting from 40. Are there other numbers she could select to increase her chance of winning? Explain your reasoning.

Item 1C: Fiji Sixes task (Fiji sixes is a lotto game)

1	2	3	4	5	6	7	8	9
10	11	12	13	14	15	16	17	18
19	20	21	22	23	24	25	26	27
28	29	30	31	32	33	34	35	36
37	38	39	40	41	42	43	44	45

If a coin is tossed twice, the outcome of the second toss does not depend on that of the first. For example, if A = {head on the first toss} and B = {head on the second toss}, B does not depend on A. A and B are said to be independent of each other. The baby problem (Item 2A) and the coin problem (Item 2B) were used to explore students' understanding of the independence concept.

Item 2A: The baby problem
The Singh family is expecting the birth of their fifth child. The first four children were girls. What is the probability that the fifth child will be a boy?

Item 2B: Coin problem

(i) If I toss this coin 20 times, what do you expect will occur?
(ii) Suppose that the first four tosses have been heads. That's four heads and no tails so far. What do you now expect from the next 16 tosses?

Analysis of the interviews indicated that the students used a variety of appropriate strategies for solving the probability problems. The data also revealed that many of the students held beliefs and used strategies based on prior knowledge which, would turn to inhibit their development of probability ideas. A simple four stage-based rubric was created for describing research results. The four categories in the are: no response, non-statistical, partial-statistical and statistical.

In this section, the main focus is on the non-statistical responses (in which students made connections with cultural experiences). Extracts from typical individual interviews are used for illustrative purposes.

4.1 Equally Likely Outcomes

Six students showed some grasp of the statistical principles underlying equal chance on the single die task (Item 1A) and two did so on the Fiji Sixes card (Item 1C). The students not only reasoned that all the outcomes had the same theoretical probability but provided correct explanations. For example, student 2 said that she did not believe that a six is hardest to get *because there is only one face with six dots and there are six faces and so the probability will be only one upon six*. In addition, student 25 not only believed that the chance of getting a six on one roll of a die was one-sixth, but was also able to provide a reason why some people believe that a six is hardest to throw. He explained that it is often the side which one needs and so one's attention is focused on the chance of a six coming up, rather than with seeing how often the six comes up compared with each of the other outcomes.

The results of this study provide information that beliefs about random generators sometimes cause students to see these common random devices in a non-normative way. Two students whose responses were classified as non-statistical on the single die task (Item 1A) believed that outcomes can be controlled by individuals. This is revealed in the following interview:

S26: It is not harder. It depends on how you throw it.
I: How do you throw so that you get a six?
S26: If you put six down and then throw you can get a six.
I: Can you try that?
S26: [Throws and gets a five].
I: Do you still think that you can make a six come?
S26: Yes.

Another student did not believe in Manoj's claims because the same force is applied when throwing a die. Luck formed an important component of these students' explanations. Two of the pupils responded on the basis of superstition, such as good luck on the single die item. For example, student 20 responded:

S20: Eh… because six is a number that starts a game eh, so if you put a six the game starts and this is luck, eh. You would be lucky and you will be able to throw a six.
I: What do you mean by lucky?
S20: Like when you playing cards, eh, so if your luck is not there, you can lose. So if you get a six you are lucky.

Student 21 thought six is not harder to throw. However, since it is a bigger number, it is difficult to get and one would be lucky to get a six.

Three of the pupils responded on the basis of superstition, such as lucky and unlucky numbers on the Fiji Sixes task (Item 1C). The students thought that a person can increase his/her chance of winning in the Fiji Sixes by selecting lucky or birthday numbers. Manifestations of the lucky number aspect are reflected in the following interview:

S14: She should have followed some other methods like I have followed, the members of the family or her lucky number.
I: What do you mean by lucky numbers?
S14: Like for me the lucky number is 18.
I: Why is 18 your lucky number?

S14: Because it is my birthday.

In addition to basing their thinking on superstitious beliefs such as luck and lucky numbers, students based their reasoning on their religious beliefs and experiences. Strong influences of religious beliefs were apparent when students were asked to comment on the advertisement regarding the sex of a baby (Item 1B). Even when challenged about how the people placing the advertisement could make money, the students could not see that roughly half the babies born would be girls and half would be boys. Anyone can expect to be right in half the number of cases just by guessing. Even if predictions are made incorrectly, some will not bother to complain anyway. Even if they did, a clear profit can be made on 50 % of the all the $20 payments sent in. The powerful nature of their religious beliefs is reflected in the response of student 17:

> As I have told you before that God creates all human beings. He is the one who decides whether a boy is born or a girl is born. Unless and until like now the . . . , have made a machine if one is pregnant and they can go there and they tell you whether the baby is a girl or a boy. But they can't tell until the baby is 8 months old. So that it means that the God created like that before we can't tell that the baby is a male or a female.

Three students referred to previous experience on the single die problem; they tended to think that it is harder to throw a six with a single die than any other score. The explanations provided by these students seemed to indicate that they remembered from their experience with board games waiting a long time for a 6 on the die, often needed to begin a game. Another student explained that since six was a bigger number it was harder to throw. The three students who referred to previous experience when commenting on the advertisement regarding the sex of the baby said that the advertisement was placed just to earn money.

Three students referred to previous experience when picking numbers from a Fiji Sixes card. For every Fiji Sixes game in the Fiji Times, there is always a sample which shows people how to play the game. In the sample, a number is crossed from each row. It seems that student 26 thought this is how numbers should be crossed, one number from each row, and experience seemed to confirm this.

S26: You have to select one number from each row.
I: Why do you say that?
S26: Because it is written in the Fiji Times example; they cross the numbers from each row. I have also seen people who play Fiji Sixes; they put numbers from each line.

4.2 Independence of Events

The baby problem (Item 2A) and the coin problem (Item 2B) were used to explore students' understanding of the independence concept. The results obtained indicated that a majority of students did not have a clear idea of the concept of independence. In three cases, when students were asked to respond to the independence tasks, they either did not respond, or gave responses which lacked information or which suggested that they could not explain. It must be noted that these responses could be indicative of reasoning based on the unpredictability bias, that is, chance is unpredictable by nature. The interview data with student 6 suggests this.

> When we toss coins or things like that we will be not sure whether we will be getting same or different all the time. So we have to make a guess.

Prior beliefs and experiences again played an important role in the thinking of those students with responses classified as non-statistical. Of the six responses classified as non-statistical on the baby problem, four related it to their religious beliefs and experiences. The students thought that one cannot make any predictions because the sex of the baby depends on God. This problem is equivalent to the coin toss task (Item 2B). When a different context is introduced, students are comfortable thinking deterministically. The religious aspect is revealed in the response of student 17 who explained:

> We cannot say that Mrs Singh is going to give birth to a boy or a girl because whatever God gives, you have to accept it.

4.3 Probability Research: A Broader Context

With respect to students' experiences, beliefs and learning, it is evident that other researchers have encountered similar factors. However, the above results suggest that in any particular context provided in the classroom students' individual learning is influenced to a certain extent by their cultural experiences. This may be problematic if students' prior experiences, values and beliefs conflict with the probabilistic concepts that teachers are trying to teach them. The findings add another dimension to educators (Kapadia 2009) who recommend that teaching in probability should be based on children's subjective ideas. Sometimes seemingly relevant experiences may get in the way of probabilistic thinking. For instance, if students believe that outcomes can be controlled by individuals or by some outside force then they need help to overcome a reluctance to predict.

In the Sharma study, the use of representativeness, equiprobability and outcome orientation was not as common as that discussed in the literature (Fischbein and Schnarch 1997; Konold 1991; Lecourte 1992). These former studies were carried out in western countries with students who would have had more cognitive experiences than the students in the Sharma study. Consequently, they would be more likely to use intuitive strategies such as representativeness and equiprobability. The students in the Sharma study had some theoretically based teaching in probability. Their ideas of chance seemed to have been influenced by rote learning, so they tended to resort to incorrect rules and procedures and previous experiences. Another explanation for this discrepancy could be that other researchers only explored students' understanding in a few contexts. In Sharma study, students' understanding in both school and out-of-school contexts was explored. Hence students were more likely to use approaches other than naive strategies. Moreover, many of the tasks that have been administered by researchers have involved forced-type responses to particular item stems. A disadvantage of forced-choice methodology is that it does not allow alternative responses to surface out. The methodology used in the Amir and William (1999) and Sharma (1997) studies attempted to explore the full range of student responses to the tasks. As a result approaches other than heuristics were evident.

A number of students could not be identified as using a consistent or pure form of reasoning but rather used a mixture of models. The explanations given by students were not consistent across similar problems. Many students who appeared to reason according to theoretical model on one problem seemed to use subjective perspective on a second problem. Different problems seemed to induce students to use different approaches. For example, student 3 drew upon the representative model on the coin toss problem (Item 2B) but for the baby problem (Item 2A) the student based her reasoning on her religious beliefs. Even when pressed about the similarity of the two problems, the student could not detect the incompatibility of her explanations. On the other hand, some students used different types of reasoning on the same problem, as illustrated by a student's response to the Fiji Sixes task.

> Yes, she could pick any out of other rows ... each row or just her favourite numbers like 2.... It is a lucky number for me. I like 12, 22, 24, 26, and 28.... It is equally likely to occur. If it has occurred in past draws, it could occur, but it is probably not likely to come up ... luck ... it has come up. Knowing most people's luck, I would choose a number that won't come out.

The student's first explanation appears to be influenced by two irrational elements: representativeness and luck. The second explanation reflects a duality of rational with irrational ideas that seemed to be influenced by beliefs. Events are seen, on the one hand, quite rationally as equally likely. On the other hand, outcomes are affected by irrational beliefs such as luck.

The interview with student 26 indicates that students may hold two rational beliefs at the same time. Events are seen, on the one hand, quite rationally as equally likely. On the other hand, outcomes are considered unpredictable by nature. In response to the question about how many heads and how many tails you would expect to get if you toss a coin 20 times, student 26 responded,

I: What do you mean by equal?
S26: 10 heads and 10 tails. But I don't know what will come.
I: What do you mean when you say "don't know"?
S26: Actually, I am not sure that when I toss the coin, if I think heads will come but tails will come.
I: Now suppose that the first four tosses have been heads. That's four heads and no tails so far. What do you now expect from the next 16 tosses?
S26: Tails 12 and heads 4.

I: Can you say why?
S26: Because I am not sure heads will come again. I have divided it equally?

As mentioned earlier, individual interviews were used in Sharma study to explore students' ideas and strategies about probability constructs. The interviews provided evidence that students often give correct answers for incorrect reasons. For example, two students thought that a six is not harder to throw (Item 1A). It would be easy to conclude from the responses that the students had well developed ideas about the equally likely concept. However, the justifications provided by the students indicate that they had no statistical explanations for their responses. None of the explanations indicated any consideration of the sample space. In some cases, students moved from appropriate strategies to inappropriate ones. One of the factors that could have made students change in this way was the students' experiences as learners at school. Students are only questioned when they give wrong answers. It seems probable that in the present research, the students interpreted the researcher's probing as an indication that something was wrong with their answers and so they quickly switched to a different strategy.

In spite of the importance of relating classroom mathematics to the real world, the findings of this section indicate that students frequently fail to connect the mathematics they learn at school with situations in which it is needed. For instance, while students could estimate theoretical probability for the die question (Item 1A), they had difficulty estimating probabilities for a real life situation (Item 1B). In fact, none of the students used theoretical or experimental interpretations for this problem. Clearly, the results support claims made by (Nasir et al. 2008) and De Abreu et al. (1992) that learning for students is situation-specific and that connecting students' everyday contexts to academic mathematics is not easy. Nasir et al. (2008) found that basketball players keep school knowledge and real world knowledge in separate compartments.

This section provides evidence of Bishop's (1988a, 1988b) assumption that formal mathematics education produces cultural conflicts between students' everyday culture and the culture of mathematics. Indeed, probability has some cultural aspects which can be differentiated in language factors, divergent everyday experiences and beliefs. The ideas behind this piece of work are important and relevant to education today. With the large numbers of minority students in mainstream classrooms with blanket probability programs it is important to actually listen to students voices to help understand what may or may not work for these students in terms of their probabilistic thinking.

5 Implications for Teachers and Research

The review of literature in the previous sections indicates that probability is not culture-free but tied to cultural practices. It may not be the same everywhere and may be related to purposes and needs of a cultural group. Additionally, the probabilistic ideas may change overtime to satisfy societal needs. This view has implications for teaching and student learning.

Whether one explains the misconceptions in probabilistic thinking by using naive strategies such as representativeness or by culturally acquired belief systems such as luck, the fact remains that students seem very susceptible to using these types of judgements. In some sense, all of these general claims seem to be valid. That is, an ordinary student has knowledge about a variety of uncertain situations. Different problems address different pieces of this knowledge. Teachers should realise that the students they teach are not *new slates* waiting to have the formal theories of probability written upon them. The students already have their own built-in biases and beliefs about probability and these cannot, as it were, be simply wiped away. If student conceptions are to be addressed in the process of teaching, then it is important for teachers of probability to become familiar with the alternative conceptions that students bring to classes.

In the Sharma study, a majority of students did not have a clear idea of the equally likely or the independence constructs. It must be acknowledged that the open-ended nature of the tasks and the lack of guidance given to students regarding what was required of them certainly influenced how students explained their understanding. The students may not have been particularly interested in these types of questions as they are not used to having to describe their reasoning in the classroom. Some students in this sample clearly had difficulty explaining explicitly about their thinking. Students who realised that it costs money to put an advertisement in the newspaper (Item 1B) had a difficult time articulating exactly how people would make money in such situations. Another reason could be that such questions do not appear in external examinations. There needs to be a focus on the language used by students and in particular on how they develop and express their ideas associated with probability.

When beginning instruction on probability, it is important for the teachers to know the individual abilities of the students. Once the level of understanding has been explored, it is crucial for teachers to teach accordingly. Teachers can accurately assess their students' understanding through individual interviews. A major disincentive for teachers is the amount of time required to interview each student. To deal with this difficulty, teachers can get an insight into all student's learning by interviewing just two or three students and generalising to the entire class.

It seems that equally likely and independence concepts are counter-intuitive and students may often hold these contradictory interpretations simultaneously. Depending on the nature of the task, this situation often generates inconsistencies in students' responses. Teachers must be aware that the level of understanding demonstrated on one context does not necessarily mean that the same understanding will be used in a different context. It appears that learning about probability constructs is a complex process and requires more emphasis and explicit teaching. They need to consider both classicist and frequentist approaches in the classroom. Such integration will enable students to construct meanings from data and build the foundation for an understanding of these ideas.

Teachers need to be sensitive to the possible effects of cultural differences, in particular as seen through religion as discussed earlier. The historical roots of probability are found in gambling. One way to bridge the gap between school mathematics and students' lives is to address issues within society that directly impact the students' lives. A good example is the study of Fiji Sixes. This game is played all over the country. The students could exchange information of the game, in particular, the popular beliefs about the game (choosing a birthday number). The next step could be modelling the odds, and the final step is the interpretation of results.

Teachers need to point out to students that there are alternative points of view. Students learn that questioning religion in any form is completely inappropriate and seen as challenging. A number of students in the study drew upon their religious beliefs when resolving their thinking. Indian cultural influences such as *God decides the sex of the baby* need to be addressed during the teaching and learning process to ensure that students construct appropriate views of learning that will promote and enhance their statistical thinking. For example, if students come to the class with the religious view that God decides the sex of a baby and the teacher is trying to teach the mathematical view that chance is blind and not controlled by prior knowledge, then how this can be done in a way that does not denigrate the first view needs to be investigated. It is important to point out to students that there are alternative points of view.

I had hoped that students would transfer probability concepts embedded in task to other contexts. Students assumed that when rolling a die all outcomes are equally likely. My hope was that they would transfer the same critical disposition to other situations involving random events. An interesting question coming out of this reason is: To what extent would students transfer their beliefs and strategies for testing the fairness of dice to other settings? Teachers would be able to address the questions with their own students and taking note of students' pattern of reasoning.

Indo-Fijian students are obedient learners. The expectations of parents and their community are that students should be respectful and polite. Clearly, given the strong values under which students work, these students may find their learning environments antithetical to their cultural expectations. This may result in an uncomfortable and unwanted ongoing experience that can hinder their learning. Bourdieu's (1984) theory of cultural capital suggests that students who have values and attitude that accord with those of their school are more likely to succeed than are students whose cultural dispositions differ. There is a worldwide phenomenon in western countries like New Zealand of increases in cultural and ethnic diversities within societies and their education systems in what has traditionally been a European dominated society. These minority ethnic groups may demonstrate a strong desire not only to engage with and succeed in the mainstream culture but also their own cultural identity. To promulgate a wider knowledge of the realities that these ethnic groups encounter, it is timely to listen to the voices of students so we can add to the growing literature on the challenges that statistics teachers face when working with students within a rapidly changing society.

A cultural approach to school mathematics will be characterized by bringing in students culture into the classroom and featuring students' cultural background in the teaching and learning of mathematics. However, giving meaning to students' cultural context in mathematics classroom is not unproblematic. What could this focus on background mean in multicultural classrooms with students from different cultures? What focus will this mean in a classroom in Fiji with students from rural and urban areas and different ethnic groups? The way in which a teacher gives meaning to an approach that focuses on the cultural background of their students hinges on many things one of which is the teachers own understanding of the concept of culture and the reasons for focusing on learners' cultural background. Teachers not only have to access, understand and accept their students' social and cultural background knowledge, they need to be able to interpret these outside realities in terms of mathematics and transform them into curriculum experiences. The question of how these approaches are translated into practice within classrooms remains. What about teachers who do not share the cultural background of their students? They will need time to work with other teachers, parents and the community in planning lessons that are relevant to their students.

5.1 Implications to Further Research

The participants in Sharma study were a fairly small non-random sample from one school. Thus, the findings—in particular, the proportions of participants who thought about probability in a particular way—may or may not generalise to the population of secondary school students as a whole. There is a need for more research with larger, more random samples with different backgrounds to determine how common these ways of thinking are in the general population.

Furthermore, as mentioned earlier, the results reported in this chapter were part of a larger study which focused on a number of areas of probability. Since there had been virtually no research focused on probability outside western countries, it was not clear when this study was conducted that the questions discussed in this chapter would be as rich and interesting as they were. Now that the cultural aspects described in this chapter have been identified as possible areas of concern, there is a need for more qualitative research focused on a deeper understanding of students' thinking about probability concepts. As mentioned earlier, Sharma study focused on students' thinking about probability before they had studied probability as a part of their senior secondary course work. There is a need for research that examines students' thinking about these concepts after they have completed their senior secondary coursework. In particular, there is a need to examine the effects of different types of instruction on students' thinking in these areas. This will help clarify cultural issues as well as ambiguous effects of teaching.

If culture is important for probability education, then one needs to consider several elements when designing tasks. First of all, researchers cannot make appropriate assessments without also having some knowledge about the range of embodied experiences in the real world of the learner. They need to choose tasks where student context knowledge base is good. The interview results show that personalisation of the context can bring in multiple interpretations of tasks and possibly different kinds of conclusions. At this point it is not clear how a learner's understanding of the context contributes to his/her interpretation of data represented in tables and graphs. There is a need to include more items using different contexts in order to explore students' conceptions of graphs and related contexts in much more depth.

While some pupils answered some questions with reference to the numerical values, others used their prior knowledge on other items. It would be interesting to see how easily (or whether) students who argued on the basis of prior knowledge on this question could be persuaded to argue purely on the basis of the numerical data. Future research could incorporate this into the interview procedure to explore this issue in more depth.

The most critical person in any learning environment is the teacher. A number of researchers have highlighted the importance of teachers' own knowledge of probability (Jones et al. 2007; Stohl 2005). With respect to culture, Greer and Mukhopadhyay (2005) warn us that probability has been introduced with minimal regard to historical and cultural contexts in spite of the fact that it is a cultural construction historically linked to religious issues. Teacher educators need to be aware of not only of students' probability knowledge but how the development of their cultural knowledge is likely to develop. Research efforts at this level are required. Through cultural studies we may gain some perspective on the forces and issues that have contributed to change in statistics. Cultural perspective will help us avoid tunnel vision about the uniqueness of the problems we face today in statistics education and suggest ways to be considered as we ponder their solutions.

5.2 Concluding Thoughts

Probability must be understood as a kind of cultural knowledge which all cultures generate but which may not necessarily look the same from one cultural setting to another. Just as all human cultures generate language, religious beliefs, rituals, house building techniques it seems all cultures are capable of generating its own probabilistic ideas. Despite this, probability curricula all over the world presumes the "correctness" of the theoretical model and ignores beliefs, values and experiences that do exist in real life situations. This presents a real dilemma which needs to be resolved if probability education is to flourish. For example, if students come to the class with the view that outcomes can be controlled by some outside force and the teacher is trying to teach the key concepts related to theoretical probability, then how this can be done in a way that does not denigrate the first view needs to be investigated. It is not adequate to consider this just as a "bias", the students, after all, may require the cultural view in some situations. Perhaps, it is important to point out to students that there are alternative points of view. It appears that probability education should indeed avoid "brainwashing" methods for alternative views are still existent (Amir and Williams 1999; Sharma 1997). It is hoped that the findings reported in this chapter will generate more interest in research with respect to cultural ideas that students possess and cultural differences. Teachers, curriculum developers and researchers need to work together to find better ways to help all students develop probabilistic thinking.

References

Abbas, T. (2002). Teacher perceptions of South Asians in Birmingham schools and colleges. *Oxford Review of Education*, *28*(4).

Amir, G., & Williams, J. (1999). Cultural influences on children's probabilistic thinking. *The Journal of Mathematical Behavior*, *18*(10), 85–107.

Averill, R., Anderson, D., Easton, H., Te Maro, P., Smith, D., & Hynds, A. (2009). Culturally response teaching of mathematics: three models from linked studies. *Journal for Research in Mathematics Education*, *40*(2), 157–186.

Barnes, M. (1998). Dealing with misconceptions about probability. *Australian Mathematics Teacher*, *54*(10), 17–20.

Batanero, C., Henry, M., & Parzysz, B. (2005). The nature of chance and probability. In G. A. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 15–37). New York: Springer.

Benn, R., & Burton, R. (1996). Mathematics: a peek into the mind of God? *Philosophy of Mathematics Education Newsletter*, *9*, 5–9.

Bishop, A. (1988a). *Mathematical enculturation: a cultural perspective on mathematics education*. Dordrecht: Kluwer Academic.

Bishop, A. J. (1988b). Mathematics education in its cultural context. *Educational Studies in Mathematics*, *19*(2), 179–191.

Bishop, A. J. (1994). Cultural conflicts in mathematics education. Developing a research agenda. *For the Learning of Mathematics*, *14*(2), 15–18.

Bourdieu, P. (1984). *Distinction: a social critique of the judgement of taste*. Cambridge: Harvard University Press.

Chiesi, F., & Primi, C. (2009). Recency effects in primary-age children and college students. *International Electronic Journal of Mathematics Education*, *4*(3), 259–274.

Clark, M. (2001). Cross-cultural issues with students from the South Pacific. *Australian Mathematics Teacher*, *57*(1), 17–20.

Cobb, P. (1994). Where is the mind? Constructivist and sociocultural perspectives on mathematical development. *Educational Researcher*, *23*(7), 13–20.

D'Ambrosio, U. (2001). What is ethnomathematics, and how can it help children in schools? *Teaching Children Mathematics*, *7*(6), 308.

De Abreu, G., Bishop, A. J., & Pompeu, G. (1992). Approaches to research into cultural conflicts in mathematics learning. In *Proceedings of the 16th PME conference*. Durham: University of New Hampshire.

Dollard, C. (2011). Pre-service elementary teachers and the fundamentals of probability. *Statistics Education Research Journal*, *10*(2), 27–47.

Falk, R., & Konold, C. (1992). The psychology of learning probability. In F. Sheldon & G. Sheldon (Eds.), *Statistics for the twenty-first century* (pp. 151–164). Washington: Mathematical Association of America.

Fijian Ministry of Education, Women, Culture, Science and Technology (1994). *Primary mathematics prescriptions*. Suva: Curriculum Development Unit.

Fischbein, E., & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for Research in Mathematics Education*, *28*, 96–105.

Fischbein, E., Nello, M. S., & Marino, M. S. (1991). Factors affecting probabilistic judgements in children and adolescents. *Educational Studies in Mathematics*, *22*, 523–549.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Schaeffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: a pre-K-12 curriculum framework*. Alexandria: American Statistical Association. Online: http://www.amstat.org/education/gaise/GAISEPreK-12_Full.pdf.

Gabriel, K. (1996). *Gambler way: Indian gaming in mythology, history, and archaeology in North America*. Boulder: Johnson Books.

Gal, I. (2002). Adults' statistical literacy: meanings, components, responsibilities. *International Statistical Review*, *70*(1), 1–24.

Gal, I. (2005). Towards "Probability Literacy" for all citizens: building blocks and instructional dilemmas. In G. A. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 39–63). New York: Springer.

Gigerenzer, G., & Edwards, A. (2003). Simple tools for understanding risks: from innumeracy to insight. *British Medical Journal*, *327*(7417), 741–744.

Greer, B. (2001). Understanding probabilistic thinking: the legacy of Efraim Fischbein. *Educational Studies in Mathematics*, *45*, 15–53.

Greer, G., & Mukhopadhyay, S. (2005). Teaching and learning the mathematization of uncertainty: historical, cultural, social and political contexts. In G. A. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 297–324). New York: Springer.

Jones, G. A., Langrall, C. W., & Mooney, E. S. (2007). Research in probability: responding to classroom realities. In F. K. Lester Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 909–955). Reston: The National Council of Teachers of Mathematics.

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

Kapadia, R. (2009). Chance encounters—20 years later fundamental ideas in teaching probability at school level. *International Electronic Journal of Mathematics Education*, *4*(3), 371–386.

Konold, C. (1991). Understanding students' beliefs about probability. In E. von Glasersfeld (Ed.), *Constructivism in mathematics education* (pp. 139–156). Dordrecht: Kluwer Academic.

Lave, J., & Wenger, E. (1991). *Situated learning and legitimate peripheral participation*. Cambridge: Cambridge University Press.

Lave, J., Murtaugh, M., & de la Rocha, O. (1984). The dialectic of grocery shopping. In B. Rogoff & J. Lave (Eds.), *Everyday cognition: its development in social context* (pp. 67–94). Cambridge: Harvard University Press.

Lubienski, S. (2002). Research, reform, and equity in US mathematics education. *Mathematical Thinking and Learning*, *4*(2/3), 103–125.

Lecoutre, M. (1992). Cognitive models and problem spaces in purely random situations. *Educational Studies in Mathematics*, *23*, 557–568.

Martin, D. (2006). Mathematics learning and participation as racialized forms of experience: African–American parents speak on the struggle for mathematics literacy. *Mathematical Thinking and Learning*, *8*, 197–229.

McKnight, C., Kallman, C., & Fisher, M. (1990). Task analyses of critical evaluations of quantitative arguments: first steps in critical interpretation of graphically presented data. In *Proceedings of annual meeting of American Educational Research Association*, Boston, MA (pp. 2–22).

Ministry of Education (2007). *The New Zealand curriculum*. Wellington: Learning Media.

Moreira, L. (2002). Mathematics education and critical consciousness'. In A. Cockburn & E. Nardi (Eds.), *Proceedings of the 26th conference of the international group for the psychology of mathematics education* (Vol. 3, pp. 369–376). Norwich: University of East Anglia.

Nasir, N. S., Hand, V., & Taylor, E. V. (2008). Culture and mathematics: boundaries between "Cultural" and "Domain" knowledge in the mathematics classroom and beyond. *Review of Research in Education*, *32*, 187–240.

Nickson, M. (1992). The culture of the mathematics classroom: an unknown quantity? In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 101–114). New York: Macmillan.

Peard, R. (1995). The effect of social background on the development of probabilistic reasoning. In A. J. Bishop (Ed.), *Regional collaboration in mathematics education* (pp. 561–570). Melbourne: Monash University.

Piaget, J., & Inhelder, B. (1975). *The origin of the idea of chance in children*. London: Routledge and Kegan Paul.

Polaki, M. V. (2002). Using instruction to identify key features of lesotho elementary students' growth in probabilistic thinking. *Mathematical Thinking and Learning*, *4*, 285–314.

Presmeg, N. (2007). The role of culture in teaching and learning mathematics. In F. K. Lester Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 435–458). Reston: The National Council of Teachers of Mathematics.

Quinn, R. J. (2003). Exploring the probabilities of "Who wants to be a millionaire?". *Teaching Statistics*, *25*(3), 81–84.
Schield, M. (2010). Assessing statistical literacy: take care. In P. Bidgood, N. Hunt, & F. Jolliffe (Eds.), *Assessment methods in statistical education: an international perspective* (pp. 133–152). New York: Wiley. Chapter 11.
Sharma, S. (1997). *Statistical ideas of high school students: some findings from Fiji*. Unpublished PhD thesis, Waikato University, Hamilton, New Zealand.
Sharma, S. (2009). Dealing with students' intuitions about probability: a social constructivist perspective. R. Averill & R. Harvey (Eds.) *Teaching secondary school mathematics and statistics: evidence-based. Practice* (Vol. 2). Wellington: NZCER Press.
Shaughnessy, J. M. (1992). Research in probability and statistics: reflections and directions. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 465–494). New York: Macmillan.
Shaughnessy, J. M., & Zawojewski, J. S. (1999). Secondary students' performance on data and chance in the 1996 NAEP. *The Mathematics Teacher*, *92*(8), 713–718.
Stohl, H. (2005). Probability in teacher education and development. In G. A. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 297–324). New York: Springer.
Strutchens, M. E., & Silver, E. A. (2000). NAEP findings regarding race/ethnicity: the students, their performance, and their classrooms. In E. A. Silver & P. A. Kenney (Eds.), *Results from the seventh mathematics assessment of the national assessment of educational progress* (pp. 45–72). Reston: National Council of Teachers of Mathematics.
Tate, W. F. (1997). Race-ethnicity, SES, gender, and language proficiency trends in mathematics achievement: an update. *Journal for Research in Mathematics Education*, *28*(6), 652–679.
Truran, K. M. (1994). Children's understandings of random generators. In C. Beesey & D. Rasmussen (Eds.), *Mathematics without limits. Proceedings of the 31st annual conference of the Mathematical Association of Victoria* (pp. 356–362). Melbourne: Mathematical Association of Victoria.
Vithal, R., & Skovsmose, O. (1997). The end of innocence: a critique of "Ethnomathematics". *Educational Studies in Mathematics*, *34*(2), 131–157.
Von Glasersfeld, E. (1993). Questions and answers about radical constructivism. In K. Tobin (Ed.), *The practice of constructivism in science education* (pp. 24–38). Washington: American Association for the Advancement of Science.
Vygotsky, L. (1978). *Mind in society: the development of higher psychological processes*. Cambridge: Harvard University Press.
Watson, J. M. (2006). *Statistical literacy at school: growth and goals*. Mahwah: Erlbaum.
Wax, M. L. (1993). How culture misdirects multiculturalism. *Anthropology and Education Quarterly*, *24*(2), 99–115.
Zaslavsky, C. (1998). Ethnomathematics and multicultural mathematics education. *Teaching Children Mathematics*, *4*(9), 502–503.
Zimmermann, G. M., & Jones, G. A. (2002). Probability simulation: what meaning does it have for high school students? *Canadian Journal of Science, Mathematic and Technology Education*, *2*(2), 221–236.

Primary School Students' Attitudes To and Beliefs About Probability

Anne Williams and Steven Nisbet

Abstract This chapter relates to the role of attitudes and beliefs in the teaching and learning of probability in schools. A study was conducted in which two Year 7 teachers in an Australian primary school and the students in their combined class participated in a teaching experiment. The study involved implementing a program of probability games and activities which aligned with both the Probabilistic Reasoning framework of Jones et al. (Stiff and Curcio (Eds.), *Developing Mathematical Reasoning in Grades K-12, 1999 Yearbook*, National Council of Teachers of Mathematics (NCTM), Reston, 1999b), and the formal Year 7 curriculum. The program was designed to improve attitudes to probability, challenge beliefs about luck, and support the learning of probability concepts. Data were collected from students and teachers with respect to attitudes, beliefs, and understanding before and after the program. It was concluded that an activity approach to the teaching of probability improved students' attitudes to and beliefs about probability, at least in the short term. Students had a greater appreciation of the relevance of probability in the world around them and their superstitions about luck lessened. There was evidence of positive links between attitudes and understanding. It was noted also that a lack of prerequisite number skills impacted on students' motivation to remain involved. At the end of the study, teachers were more confident and enthusiastic about teaching probability in the future.

1 Introduction

The affective domain in mathematics has been given much focus in research over the years. Drawing upon McLeod's (1992) work on attitudes, DeBellis and Goldin (2006) and Grootenboer and Hemmings (2007) have characterized the affective area as including beliefs, values, attitudes, and emotions. Some of these overlap and interact with each other. In particular, the role of attitudes has occupied an important place in the study of teaching and learning because of the perceived links between

A. Williams (✉) · S. Nisbet
Griffith University, Brisbane, Australia
e-mail: amawwill@bigpond.net.au

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_36,

attitude and learning outcomes (Dweck 1986; Estrada 2002). With respect to probability or 'chance', Gal (2005) affirms that attitudes "play a key role in how people think about probabilistic information or act in situations that involve chance and uncertainty" (p. 45). Beliefs about probability have been highlighted because of the misconceptions, inconsistencies and difficulties related to probabilistic reasoning (see Shaughnessy 1992). Given this research, this paper investigates the attitudes and beliefs of primary students as they engage in probability games and activities.

1.1 Attitudes

It is generally accepted that attitudes are acquired, and can be modified, through learning and experiences over time; and that attitudes can be inferred through behaviour (Krosnick et al. 2005). Emotions are less stable than attitudes, but repeated emotions stabilise into attitude (McLeod 1992). The relationship between attitudes and behaviour can be explained in the theory of personal action (Fishbein and Ajzen 1975; Ajzen and Fishbein 2000) (Fig. 1), which suggests that attitudes influence intentions, which influence behaviour. Behaviour then leads to personal experiences which in turn have an effect on attitudes.

Applying the theory of personal action to the learning of mathematics, it is possible to hypothesise both a positive and negative cycle (Nisbet 2006). In the positive cycle (Fig. 2), a student with positive attitudes to mathematics has the intention to do well, hence exhibits positive behaviours, such as attending to the task, working hard, or seeking help, which lead to successful experiences. In turn, this success produces positive attitudes, such as increased motivation, and the cycle continues. When students are motivated, they attend to instruction, strive for meaning and persevere when difficulties arise, thus strengthening the positive cycle (Aiken 1970; Holmes 1990).

In the negative cycle (Fig. 3), a student with negative attitudes to mathematics often exhibits disinterest (Cathcart et al. 2006) and less inclination to try in class (Bernstein et al. 1979). Perceptions about the value of a task can also affect motivation (Woolfolk and Margetts 2007), and anxiety about mathematics can reduce school achievement (Covington and Omelich 1987). These factors may work against

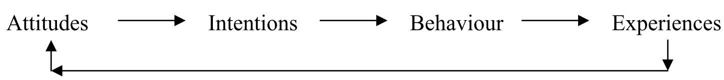


Fig. 1 The attitudes–behaviour cycle

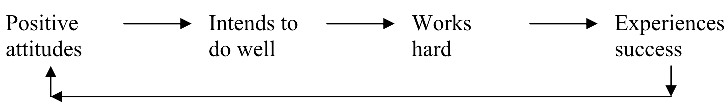


Fig. 2 The positive attitudes–behaviour cycle

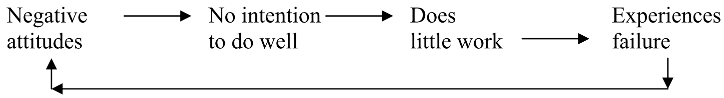


Fig. 3 The negative attitudes–behaviour cycle

the student in the negative cycle. As a result of negative attitudes, less work is done in class, and failure is experienced. In turn, this produces negative attitudes, and the cycle continues. Examining nine different studies, Aiken (1970) noted that consistent failure in mathematics results in frustration, anxiety, a loss of self-confidence and further development of "dislike and hostility toward the subject" (p. 587).

Thus, the attitudes–behaviour cycle in mathematics is quite complex, and can involve aspects of self-confidence, interest in the domain, anxiety, motivation, and perceptions of value. In fact, these aspects constitute the basic attitude dimensions consistently used in many attitude studies. For example, Fennema and Sherman (1976) proposed a set of scales for the measurement of attitudes in secondary students—attitude toward success in mathematics, confidence, anxiety, effectance motivation (enjoyment), usefulness of mathematics, mathematics as a male domain, and perceptions of mother's, father's and teacher's attitudes toward learning mathematics. More recently, Tapia and Marsh (2004) identified four attitude factors from their study of secondary students' attitudes to mathematics, namely self confidence, value of mathematics, enjoyment and motivation. The present study draws upon these studies in using the constructs of enjoyment and interest, confidence, anxiety, perceptions of usefulness, and motivation, and focuses on the attitudes of students in Year 7—the final year of primary school in Queensland, Australia.

Research studies on school students' attitudes to probability are few, particularly at the primary school level. This study addresses this shortage. In the related field of statistics, Gal and Ginsburg (1994) noted that difficulties with learning statistics were due to non-cognitive factors such as negative attitudes and inappropriate beliefs, which impeded students' development of statistical intuitions and applications beyond the classroom. Williams (1992) reported that tertiary students with positive attitudes to statistics were more likely to report positive experiences and success with school mathematics, enjoyment and a sense of challenge with statistics, an appreciation of its usefulness, and a need for understanding by developing appropriate strategies. Onwuegbuzie (2000) identified 'statistics anxiety' as a correlating factor in graduate students' academic achievement. Thus the literature in the statistical domain supports the relationship between success and positive attitudes.

Breaking the negative cycle, by promoting attitudes associated with successful experiences such as interest, motivation and confidence, would seem to be an admirable aim of classroom instruction. However, students' ways of connecting with a domain are linked to their ways of participating within the mathematics classroom (Gresalfi and Cobb 2006). It follows that there should be links between teaching methods, teacher knowledge, and the moulding of attitudes and beliefs in mathematics (Wilkins and Ma 2003; Zan and Martino 2007). Working against making these links is the fact that 'chance' has been found to be a topic that teachers do not like, and feel unprepared to teach (Watson 2001; Carlson and Doerr 2002). Many teachers lack domain-specific preparation and knowledge, and there is a lack of specific training and support in the area (Batanero et al. 2004). According to Joliffe (2005), some teachers lack confidence when teaching topics involving numeracy, and this could affect many aspects of their teaching. Joliffe continues that "Research into attitudes and beliefs could be useful in discovering and helping to overcome this problem" (p. 340). A teaching intervention that aims for student success and increased teacher knowledge and confidence through positive experiences with probability is therefore appropriate.

In this study, games and activities were used as a means of providing such experiences. Games have a positive influence on the affective component of learning situations while still contributing to the development of knowledge (Booker 2000). They can raise levels of students' interest and motivation (Bragg 2007), and encourage logico-mathematical thinking (Kamii and Rummelsburg 2008). The conduct of suitable games and activities in the classroom may also assist in improving teachers' confidence with the topic of probability.

Many writers in the area of motivation distinguish between intrinsic and extrinsic motivation. For example, Jones et al. (2011) note that it makes sense for teachers to intrinsically motivate students because of the many positive outcomes (e.g. good quality learning, creativity); extrinsic motivation has many negative outcomes (e.g. student anxiety, less flexible thinking). They claim that using suitable manipulatives in mathematics lessons in ways that (i) provide for substantial student autonomy, (ii) provide opportunities for success, and (iii) allow students to work together in groups with a sense of belonging, promotes greater intrinsic motivation. It can be argued that the use of games in probability lessons (with the inherent manipulatives such as dice, spinners, coins, etc.) has the potential to satisfy all three criteria, and lead to increases in levels of intrinsic motivation.

A substantial body of research confirms that many teaching strategies designed to increase student motivation at school, particularly in mathematics, also improve learning outcomes (Stipek et al. 1998; Bobis et al. 2011). This conclusion along with that of Jones et al. (2011) would imply that using games in probability lessons leads to improvements in attitudes, beliefs and understanding.

In this study, we investigated the use of probability games and activities as a strategy to improve primary (Year 7) students' attitudes to the topic of probability and improve understanding of probability concepts. At the same time we investigated ways of assisting their teachers to gain confidence and improve their sense of preparedness to teach probability using the activity approach.

1.2 Beliefs

Beliefs are internal knowledge that directs actions and learning (Lester 2002). Beliefs are seen as less intensive, more stable, and more cognitive than attitudes (McLeod 1992). As with attitudes, beliefs are influenced by classroom experiences

(Yackel and Rasmussen 2002). According to Grootenboer and Hemmings (2007), beliefs and attitudes have overlapping elements. For example, confidence in doing mathematics, an attitude, and belief in one's ability to do mathematics may influence each other (McLeod 1992).

Early research in probability (e.g. Tversky and Kahneman 1973) showed that beliefs can influence judgements about the likelihood of events, such as predicting an outcome based on the recall of personal experiences (*availability* heuristic). More recently, Truran (1995) found that many Year 3 children held erroneous beliefs about outcomes (e.g. holding the die in a certain way, praying for the desired outcome, changing the shape of the random generator). Fischbein and Gazit (1984) and Konold (1991) noted that erroneous beliefs can still persist regardless of opposing empirical evidence, but instruction highlighting the discrepancies could be advantageous. In fact, Jones et al. (1999a, 1999b) found that probabilistic thinking in Year 3 did improve significantly with instruction. Watson's longitudinal study of students starting Years 3, 6 and 9 (Watson et al. 2004; Watson et al. 1995, 1997) recommended that instruction highlight context, rather than personal experiences, to develop belief structures, which she had concluded were under-developed. The notions of luck and superstition also enter into studies on probability beliefs. Watson's research found that probability outcomes explained by luck were based on non-mathematical beliefs, whereas quality explanations of outcomes involved mathematics (e.g. calculations, ratio comparisons). Amir and Williams (1999) showed that 11 to 12 year-old children, particularly those with lower ability, demonstrated a "high degree of superstition" (p. 94). Given the above research, instruction to promote understanding and modify any existing erroneous beliefs would be beneficial. In this study we also investigated the effect that students' engagement with chance games and activities had on their beliefs of about luck—specifically so-called lucky people and lucky strategies.

In summary, it was hoped in this study that the teaching intervention utilising games and activities would enhance students' probability attitudes, modify any erroneous beliefs, prevent the negative attitudes-behaviour cycle from developing, and increase students' understanding of probability concepts. It was also hoped that the teaching intervention would better equip and motivate teachers to teach probability. For this paper, the questions posed are: Have the games and activities, purposefully selected to develop the key probability concepts, brought about changes in students' and teachers' attitudes and beliefs about probability? What is the nature of the changes and their effect on behaviour? Is there evidence of any changes in students' thinking and performance in 'chance'?

2 Methodology

2.1 Participants

The participants were two experienced female Year 7 teachers (designated in this study as Teacher A and Teacher B) working cooperatively in a suburban government school in the state of Queensland, Australia, along with their 58 students in the combined class (53 % boys, 47 % girls).

2.2 Method

The study was a small-scale teaching experiment (see Steffe and Thompson 2000), which permitted cycles of observation, interaction with students, reflection, and planning. Each lesson used two probability games or activities, and was taught by one of the researchers. A day or two later, the lesson was repeated by the teachers in their individual classes. There were three such cycles of lessons spread over a two-week period.

2.3 Games/Activities

The four probability games and two probability activities were deliberately selected to reflect and develop the theoretical constructs of the Probabilistic Reasoning Framework (Jones et al. 1999a, 1999b), namely randomness, likelihood, sample space, experimental and theoretical probability, and independence. The games/activities were aligned with relevant curriculum content, and were designed to motivate and challenge students' thinking in terms of predicting and explaining possible and actual outcomes. For example, 'Get Your M&Ms' was a game played in pairs, where children placed their 12 M&Ms on a game board containing numbers 1 to 12. If, on the roll of two dice, the total matched a board number with an M&M on it, the M&M could be removed. Winning involved removing all 12 M&Ms before the other player. Students soon learned that some dice totals occurred more frequently than others, and the sum of 'one' did not occur. Addition and frequency tables prompted students to consider a good strategy for winning the next game. This game provided opportunity for the discussion of all of the probability constructs in the Probabilistic Reasoning Framework.

The complete list of activities used in the project is as follows. For more detail, see Nisbet and Williams (2009).

1. 'Greedy Pig': Dice game. Students take risks on the likelihood of a 'poison' number (e.g. a '2') being rolled by the teacher.
2. 'Get your M&Ms': [See paragraph above.]
3. 'Dicey Differences': Two-dice game. Player A wins a point if the difference is 0, 1, or 2; Player B wins a point if the difference is 3, 4, or 5.
4. 'Multiplication Bingo': Game. Each student inserts 16 different numbers on a 4×4 bingo board, hoping to match cards selected randomly from a pack containing single-digit multiplication facts called by the teacher.
5. 'Peg Combo': Activity. Each student draws one peg (and then later, two pegs) from a brown paper bag containing two red and two blue pegs. Actual and expected results for the whole class are compared.

6. 'Rolling Dice': Activity. Students predict the results from 60 rolls of a die, then make observations while conducting three sets of 20 rolls. Experimental and predicted data are compared.

2.4 Data Collection

During the project, data on attitudes to and beliefs about 'chance' were collected through student surveys (pre-test, post-test), teacher interviews (pre-test, post-test), teachers' journals, and end-of-study interviews with selected students. An eclectic approach to gathering data in the statistical domain had been recommended by Gal et al. (1997). Pre- and post-test data were also collected on students' cognitive performance on 'chance' concepts, but only quantitative totals of the test items were used in this study.

The student survey of attitudes and beliefs consisted of 14 items relating to attitudes to and beliefs about 'chance'. As discussed previously, the 'attitude' constructs on the survey were: (i) enjoyment and interest, (ii) confidence, (iii) perception of usefulness of 'chance', (iv) anxiety, and (v) motivation. The two 'belief' constructs were: (i) lucky people, and (ii) lucky strategies. This survey was designed by the researchers, because in general, previous attitude surveys neither related to probability specifically, nor were appropriate for primary school students. The focus here was on observing any changes in attitudes and beliefs. The test–retest reliability coefficient for total attitude scores was found to be 0.814, with the time interval between the two tests being three weeks. Correlations between total scores on matching items (same construct) for the pre- and post-tests were found to be significant at the 0.05 level of significance. Thus there was evidence of stability between the two administrations of the survey. Half of the items were worded positively, half negatively. Responses to the items were obtained on a five-point Likert scale, from 1 (Disagree strongly) to 5 (Agree strongly), and students were asked to provide reasons for each response. Reasons were categorised (see Table 2). To analyse and interpret results (see Table 1), all responses were later scored and recorded from most negative (1) to most positive (5).

The teacher interview questions related to the student survey of attitudes and beliefs constructs, but also teaching experiences (e.g. To what extent do you enjoy teaching about 'chance'? Please explain your answer). During the project, teachers were requested to keep journals, recording their observations of students, any changes over time, and their reflections on the activities and the teaching/learning processes.

At the end of the study, individual interviews were conducted with six students, designated as L1, L2 (low ability), A1, A2 (average ability), and H1, H2 (high ability). These students were present for the pre-tests and almost all of the activities. Each group of two was randomly selected from groups of students designated by the teachers to have low, average, or high mathematics ability. The interview included questions on students overall reflections about the 'chance' lessons, their levels of enjoyment of each of the activities, and any difficulties they encountered.

3 Results

Results in this section are analysed in the following ways:

- Pre–post comparison of student survey of attitudes and beliefs responses
- Analysis of extended responses on the student survey of attitudes and beliefs
- Teachers' perspectives
- Case study—six students
- Additional quantitative analysis

3.1 Pre–Post Comparison of Student Survey Responses

It was hoped that teaching 'chance' through games and activities would improve students' attitudes to and beliefs about 'chance'. A pre–post comparison of the survey data was conducted using paired t-tests in order to discern any significant changes.

At the end of the study, Table 1 indicates that students' *attitudes* had improved significantly in four areas:

- Students expressed more enjoyment when learning about 'chance' (Item 1);
- Students were less anxious about doing 'chance' in class (Item 4);
- Students were more motivated to learn more about 'chance' (Item 5); and
- Students thought that there were many more uses of 'chance' to learn about (Item 11).

There was also a significant change in students' *beliefs* about luck.

- Students were less inclined to believe that some people are born lucky (Item 6).

Furthermore, total scores improved significantly (see last line of Table 1).

- There was an overall improvement in students' general attitudes/beliefs about 'chance'.

Comparison of the effect sizes on those items that were significant with those items that were not revealed that the former were all higher than the latter. For example, the effect sizes on significant items ranged from 0.28 (Item 1) to 0.53 (Item 4), while those on non-significant items ranged from −0.20 (Item 8) to 0.21 (Item 11). Hence, effect sizes seem to support the findings.

3.2 Analysis of Extended Responses on the Student Surveys

Table 2 records the major ways in which students supported their initial response ("Disagree strongly" to "Agree strongly") on each item on the survey. Each extended response was categorised. For example, on Item 3 (usefulness), responses such as *It is because everything or nearly everything is about chance* and *because*

Table 1 Pre–post comparison of student survey items

Item	Statement	Pre-test mean response	Post-test mean response	df	Mean diff (post-pre)	sd	*t* value for paired t-test	*p* value (one-tailed test)	Effect size (mean-diff/sd)
1	I enjoy learning about 'chance'.	3.88	4.08	47	0.208	0.743	1.944	0.029*	0.28
2	I am willing to have a go when we learn about 'chance' in class.	4.17	4.25	47	0.083	0.647	0.893	0.19	0.13
3	I don't think 'chance' is a useful topic to learn about.	3.98	3.98	47	0	0.968	0	0.5	0.00
4	I get worried when we do 'chance' in class.	3.92	4.33	47	0.417	0.794	3.633	0.0003*	0.53
5	I would like to learn more about 'chance' in class.	3.83	4.02	47	0.187	0.673	1.929	0.03*	0.28
6	Some people are born lucky.	3.45	3.87	46	0.426	1.175	2.483	0.008*	0.36
7	Crossing your fingers does not improve your luck at all.	3.85	3.64	46	−0.21	1.545	0.944	0.18	−0.14
8	'Chance' is not a very interesting topic to learn about in class.	4.06	3.89	46	−0.17	0.868	−1.345	0.09	−0.20
9	I would rather not have to try when we learn about 'chance' in class.	3.96	4.11	46	0.149	0.932	1.096	0.14	0.16
10	'Chance' does not worry me at all in class.	3.98	4.02	46	0.043	0.908	0.321	0.37	0.05
11	There are many uses of 'chance' to learn about.	3.83	4.11	46	0.277	0.926	2.049	0.023*	0.30
12	I hope we don't have to learn about 'chance' again next year.	3.89	4.06	46	0.170	0.816	1.430	0.08	0.21

Table 1 (Continued)

Item	Statement	Pre-test mean response	Post-test mean response	df	Mean diff (post-pre)	sd	*t* value for paired t-test	*p* value (one-tailed test)	Effect size (mean-diff/sd)
13	You can improve your chances of winning by using lucky numbers, birthdays and house numbers.	3.7	3.74	46	0.043	0.833	0.350	0.36	0.05
14	No one has any more luck than anyone else.	3.7	3.83	46	0.128	1.393	0.628	0.27	0.09
	All survey items	54.23	56.02	46	1.787	4.222	2.902	0.003*	0.42

*Significant at the 0.05 level of significance

we don't really need it in later life would both have a category of "relevance", the first in a positive sense, the second in a negative sense. A response such as *Knowledge about chance would help in tests and problem-solving* would be classified as a positive response relating to "learning" about 'chance'. On Item 6 (lucky people), a response such as *because some people get lucky all the time* would have a category of "superstitious" based on "experience", while a response such as *it's only a saying*, would have a category of "not superstitious" with a reason "other" than one based on experience, or an understanding of 'chance'.

On the *attitude* items (Items 1 to 5, 8 to 12), several observations can be made:

- The main categories were stated in a positive sense.
- Extended responses focused on relevance, challenge, interest, easiness, fun and learning aspects. The latter two were more prevalent.
- For the majority of items, the most prevalent category did not change from pre- to post-test (exceptions were Items 5 and 10).

On the *belief* items (Items 6, 7, 13, and 14), additional observations can be made:

- Extended responses for the majority of students were of a "non-superstitious" nature.
- The percentage of students providing "non-superstitious" responses increased between the pre-and post-test, and categories of responses varied.
- The percentage of students providing "superstitious" responses decreased between the pre-and post-test.
- The percentage of extended responses demonstrating some understanding of 'chance' increased on the post-test (see Items 13, 14).

Table 2 Major two categories of students' extended responses, and percentage of students by category

Item	Statement	Major categories*	Pretest %	Post-test %
1	I enjoy learning about 'chance'.	+Fun	22	56
		+L'ing	17	
		+Relev		14
2	I am willing to have a go when we learn about 'chance' in class.	+L'ing	24	42
		+Fun	22	22
3	I don't think 'chance' is a useful topic to learn about.	+Relev	56	67
		+L'ing	16	15
4	I get worried when we do 'chance' in class.	+Easy	29	33
		+Other	19	22
5	I would like to learn more about 'chance' in class.	+Fun	17	46
		+L'ing	38	37
6	Some people are born lucky.	NS–Other	51	66
		S–Other	18	(4)
		NS-Exp'ce		13
7	Crossing your fingers does not improve your luck at all.	NS–Other	57	57
		S–Other	18	(2)
		NS-Exp'ce		17
8	'Chance' is not a very interesting topic to learn about in class.	+Int'ing	45	50
		+Fun	38	25
9	I would rather not have to try when we learn about 'chance' in class.	+Chall	23	45
		+L'ing	23	(13)
		+Fun		23
10	'Chance' does not worry me at all in class.	+Easy	32	23
		+Other	27	26
11	There are many uses of 'chance' to learn about.	+Relev	29	64
		+Exp'ce	14	(7)
		+L'ing		11
12	I hope we don't have to learn about 'chance' again next year.	+L'ing	36	51
		+Fun	26	31
13	You can improve your chances of winning by using lucky numbers, birthdays and house numbers.	NS-Other	66	33
		S-Other	17	(14)
		NS-Undch		38
14	No one has any more luck than anyone else.	NS-Undch	55	59
		NS-Other	16	14
		S-Exp'ce		14

*L'ing, Learning; Relev, Relevance; Int'ing, Interesting; Chall, Challenging; NS, Not superstitious; S, Superstitious; Exp'ce, Experience; Undch, Understanding of 'chance'; NS/S-Other, response not associated with experience nor an understanding of 'chance'; Other, response not associated with interest, fun, motivation, challenge, curiosity, easiness or acceptance; +, positive sense; –, negative sense

3.3 Teachers' Perspectives

On Their Own Attitudes and Beliefs At the beginning of the project, Teacher A expressed concern about her lack of confidence and understanding of 'chance'. These diminished her capacity to extend children beyond having fun, recording the results of coin tosses, or classifying events. She felt inadequate to take the confusion out of the topic for her children. She was "a strong believer in learning through fun, activities that are hands-on and motivating, and especially group work." She wanted to know how to challenge her students and understand strategies to improve predictions in 'chance' situations. At the end of the project, she now felt more confident about teaching 'chance' because her own understanding was developing. For example, she had learnt about the *chance tree* (tree diagram) which was "fantastic"; had a set of new strategies and activities to use; and could relate 'chance' to her everyday life. The study confirmed her non-belief in lucky strategies. Now, she felt that it was important to have "a lot more discussion and de-briefing about the games".

At the beginning of the project, Teacher B stated that she loved, and was confident in, teaching mathematics. For her, visualising and thinking were valuable skills. 'Chance' was important because unlike other areas of mathematics, "children's answers can vary depending on how well they conceive what is happening." "Lucky" people possessed a positive outlook which exuded confidence, acceptance, and determination; they are more likely to succeed, hence perceived as "lucky". At the end of the project, she expressed satisfaction that the children were very involved, and that obtaining a correct answer was less important than thinking. She felt that her views on teaching 'chance' had widened, and that she would try different things in the future, with more questions and discussion. She would no longer assume that doing means learning. She now saw 'chance' as very relevant in helping people to make wiser choices in life.

Hence, for both teachers, the project was seen as successful in providing them with a different perspective on teaching and learning 'chance'. They now had new knowledge, confidence, additional activities and strategies, and a greater appreciation of the uses of 'chance'. This would no doubt impact on their future approaches to teaching 'chance' at this level. Their beliefs about luck had not changed greatly.

On Their Students' Attitudes Teacher A noted that the children were "engaged from the start" in activities such as 'Greedy Pig' and 'Get Your M&Ms', and were really keen to play them again in the follow up sessions, which were "hilarious". She noted that children were "confident" when playing 'Dicey Differences', and keen and competitive with 'Multiplication Bingo'. However, while those who had a good grasp of their number facts could respond instantly and were on task, the weaker students "fell behind—got frustrated—and anxious." "It was also a fun way to get (students) to work on (number facts)." Teacher A observed that in some repeat sessions some children were not keen on writing another response, possibly because they could not say anything new. Teacher A was happy that some children (the "thinkers") spent "considerable time writing and responding to the questions", and that "children's feelings & comments (were) sought and valued." She noted that the children loved: playing in pairs, vying against each other, the "hands on" aspect; they were "all totally engaged & motivated", and having fun throughout. Some students were predicting, trying new strategies, and playing an extra game if they finished early. In the repeat session with 'Greedy Pig', there was discussion such as on the way the die was thrown, the possibility that it was weighted, and the type of die used. These ideas were consistent with the findings of Truran (1995) and Amir and Williams (1999). However, despite the positive aspects of the project, Teacher A expressed concern that "Many of the children… (were) happy to do the 'fun part' but are not bothered to think about probability without being prompted or prodded by the teachers."

Teacher B stated that children enjoyed the 'chance' activities because they were "hands-on." Further, the activity approach gave all levels of student "something to enjoy". She observed that during the repeat lessons, students appeared more confident with the activities and "wanted to be independent, and do it at their own pace. They just get into it the second time." Teacher B noted that all students participated well, and the games were played with enthusiasm. For example, in 'Multiplication Bingo', "there was joy at being able to win a point, sighs at the game being over." However, she suggested that because they had been asked to think, some students were hesitant in answering questions. Others were "a little bit scared" about writing a response. She argued that students would see mathematics as "a right or wrong" subject where "you tick or cross"; some students would be thinking, "Have I got to think of something really intelligent?"

Hence the teachers' observations were that the nature of the activities ensured interest and enjoyment, as well as motivation to remain involved throughout the project. Children seemed more confident during the repeat sessions, and particularly when their background knowledge was sound. Many were challenged to ask questions, or consider factors and strategies that could influence results. Some anxiety was evident when students were asked to provide oral or written rationalisations, especially those less mathematically able.

3.4 Case Study—Six Students

The six interview students were asked to comment on what they thought of the 'chance' lessons overall, and to rate the games and activities on how much they enjoyed each of them on a scale of 1 (not at all) to 5 (very much). They were also asked to comment on why they enjoyed the games and activities, what they learned from them, and any difficulties encountered.

The general consensus was that the games were easy, fun to play, interesting and enjoyable. Across all ability levels, students showed a preference for the activity approach to learning. Reasons given included:

- Novelty, e.g. "so you can play games to learn about chance, instead of just writing stuff" (H2); and

Fig. 4 Students' enjoyment ratings according to students' ability level

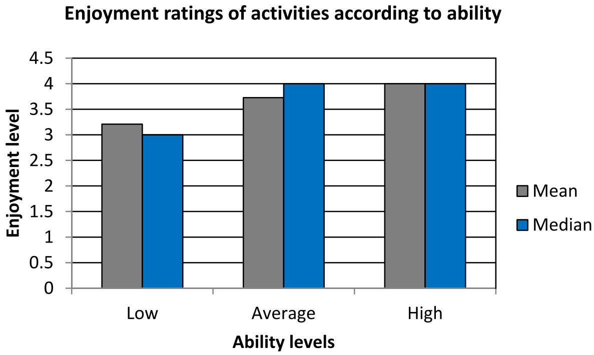
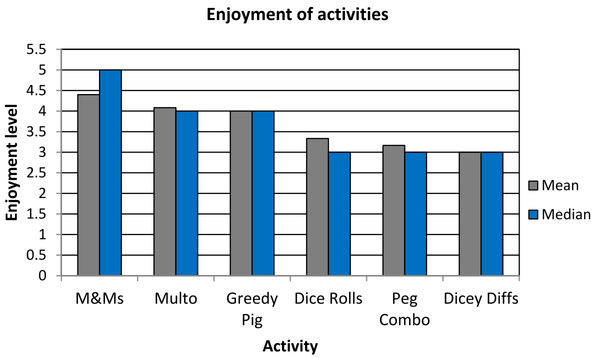


Fig. 5 Students' enjoyment ratings according to game/activity



- Improved perceptions about luck, e.g. "(Having a strategy) is a faster way of winning a game" (H1).

Hence the games/activities were enjoyable and motivating, a preferred way of learning, and they demoted the role of luck in winning certain games.

In Fig. 4, enjoyment ratings for the games/activities appear to vary according to students' ability level. However, the Kruskal–Wallis test indicated no significant differences in the medians of enjoyment ratings with respect to ability level ($p = 0.069$).

In Fig. 5, enjoyment ratings on each game/activity are graphed.

According to the interview students, the most enjoyable game was 'Get your M&Ms' (mean rating = 4.4; median rating = 5). All six students thought it was fun. Reasons given were:

- Motivation (e.g. "because if you won you got to eat some M&Ms" (A2));
- Enjoyment of playing games with a friend (L1).

Most students quickly caught on to a valid strategy, namely, putting more M&Ms towards the middle of the game board: "you can never put a one on" (L1). A2 tried different ways of throwing the dice, and L2 said she was frustrated when her chosen numbers were not thrown. Hence the game provided enjoyment, motivation, and at times some anxiety; students were challenged to think or try something different, and in this context, the students did develop relevant understandings of 'chance'.

The next most enjoyable game was 'Multiplication Bingo' (mean rating = 4.1; median rating = 4). The reasons given by the students referred to:

- The element of competition (e.g. "you feel good when you win" (H2));
- The challenge and motivation to learn (e.g. "because I knew my multiplication tables, and I could figure out which numbers would be coming up" (A1)).

L1 expressed frustration that his seemingly good combination of numbers did not help him win. In this game, the students again quickly found a valid strategy, namely, using numbers with more factors. Thus this game provided the motivation to extend learning through student involvement and enjoyment, but the randomness of the numbers selected was seen to negate the impact of good strategies at times.

The third most enjoyable game was 'Greedy Pig' (mean rating = 4; median rating = 4). Reasons for enjoying the game included:

- Surprise and pleasure (e.g. "You never know when two would come up… it's very sudden" (L1));
- Inclusion (e.g. "everyone participated and yeah it was really enjoyable because it was like a game, and you don't realise that you actually learn from a game" (H1));
- Learning (e.g. "it teaches us to take or not to take risks" (H2)).

Several strategies for winning were developed, such as "I waited until (I had a total of) 15 to 20, or 10" (A2). Thus the pleasure, participation and surprise elements were such that students did not realise they were learning from playing the game; strategy development was an outcome.

'Dicey Differences' was the activity least enjoyed by the students (mean rating = 3; median rating = 3). The most common reason given for this was the unfairness of the game. Other comments indicated confusion about how to play the game and how to record results. However, most students learnt that "some (numbers) have more chance of winning and some have more frequencies" (A1). Three of the six students could devise rules for a fair game. Thus the perceived instructional impediments and the intrinsic unfairness of the game appeared to act as demotivators to enjoyment and completion of the learning process.

'Peg Combo' was next least popular activity (mean rating = 3.17; median rating = 3). The main reason was the nature of the activity (e.g. "it was not really as fun as the others" (L1)). However, H1 was excited because the tree diagram explained things for him: "I find the strategy quite useful, because in lots of math sessions they do kind of ask you about variations, what variations you could (have)." Most students seemed to understand that mixed colours occur more frequently than individual colours. Hence, despite feeling that this game was less enjoyable and interesting than other games, most students gained knowledge about 'chance' from playing it.

The other activity in the least-enjoyable group was 'Rolling Dice' (mean rating of 3.33; median rating = 3). Reasons for ratings were mixed, including:

- Boredom with repetition (e.g. it was just really rolling dice" (A1));
- Unpredictability (e.g. "you never know what's going to happen" (A2));
- Challenge (e.g. "well, to my surprise, I had no idea what was going to happen to do with chance, so I just put ten in each column... because 60 times, and six numbers, so you divide 60 by six numbers which is ten" (H1)).

Three of the six students (H1, H2, and A1) realised that the frequency for each number on the die would even out after many trials. The others stated that it was difficult to predict frequencies because of the randomness of the outcome on the toss of a die: They did not notice the pattern in the percentages they calculated. Hence, lack of enjoyment, interest, and sense of challenge were experienced, with mixed cognitive results. Reflection on the design and method of playing this game would be advisable for future use.

Thus the case study results lend support to using games and activities such as these to improve attitudes to learning 'chance'. Students enjoyed the games and activities because they were fun to play, or a novel way of learning. The most enjoyed activities motivated students to learn and try different strategies; challenged them to try to make sense of the outcomes through reasoning, experimentation and discussion; and gave them the confidence to accept losing despite having a sound strategy. At times, learning seemed to be an unexpected outcome from participating in an enjoyable activity. For the least enjoyed games/activities, it was often their nature (e.g. repetitious, unpredictable) that appeared to act as a demotivator to enjoyment and challenge, yet not necessarily a barrier to learning.

3.5 Additional Quantitative Analysis

Quantitative analysis of cognitive performance on test items revealed a statistically significant overall improvement in performance from the pre- to the post-tests ($t = 2.5$, one-tailed $p = 0.008$).

Quantitative analysis of the relationship between attitude and cognitive performance showed the following results:

(a) There was no significant correlation between total student survey score and total cognitive test score on the pre-tests ($r = 0.18$, 2-tailed $p = 0.192$).
(b) There was a low significant correlation between total student survey score and total cognitive test score on the post-tests ($r = 0.31$, 2-tailed $p = 0.03$).

4 Discussion and Conclusion

The intent of this study was to use a teaching intervention with games and activities to promote a positive cycle of attitudes, behaviour and experiences, and lessen the potential for a negative cycle to form. Including teachers in the teaching/learning cycle was expected to improve their understanding and motivation to teach probability. Earlier in this paper, the following questions were posed:

- Have the games and activities, purposefully selected to develop key probability concepts, brought about changes in students' and teachers' attitudes to and beliefs about probability?
- What is the nature of the changes?
- Is there evidence of changes in students' thinking and performance in 'chance'?

The results are now collated and discussed.

Students' overall attitude/belief scores on the survey and their total cognitive test scores both increased significantly between the pre- and post-tests. Also, there was a low significant correlation between total survey score and total cognitive test score on the post-tests, which did not exist with the pre-tests. Hence the quantitative analysis suggests positive changes in both attitudes/beliefs and performance, and some relationship between the two. Other findings are discussed in terms of the survey constructs.

Enjoyment and Interest On the student survey of attitudes and beliefs, the t-tests indicated significantly more enjoyment in learning 'chance' at the end of the study (Item 1). There was a non-significant increase in the perception that 'chance' was an interesting topic to learn about (Item 8). The extended responses on these items were positive, focussing on fun, interest, learning, and the relevance of 'chance'. Collectively, the teachers expressed satisfaction that children with different levels of ability could enjoy many aspects of the activities, namely, the "hands-on" aspect, collaboration, competition, involvement, thinking, and lack of emphasis on a "correct" answer. The teachers were happy to have increased their own knowledge of probability, and looked forward to trying new things with their classes.

The positive comments of the interview students were similar to those of their teachers. Interview students also appreciated the variation from normal mathematics lessons, the motivating, challenging and risk-taking nature of the games, and strategy development. Thus, the games/activities stimulated enjoyment, that in turn led to increased motivation, participation, and thinking. Thus enjoyment fostered Nisbet's (2006) positive-attitudes–behaviour cycle.

The comments for the least-enjoyed activities related to unfairness, confusion in the rules of play, their 'boring' nature, and repetition. In these cases, interest waned. Teacher A noted that children were often "happy to do the fun part but not prepared to think about any other probability without being prompted or prodded by the teacher".

Randel et al. (1992) examined 68 studies on the effectiveness of educational games in a range of school subjects. The highest interest in learning with games was in mathematics. Randel et al.'s findings implied that sustaining interest and involvement through games increased the potential for cognitive structures to be altered to include new knowledge. In the current study, interest could be maintained by changing the activities that were boring and repetitive, or by supplementing them with computer simulations (see Lee and Lee 2009; Sánchez 2002; Wood 2005).

Confidence On the student survey of attitudes and beliefs, the t-tests showed non-significant increases in confidence between pre-test and post-test means (Items

2 and 9). The extended responses focused on enjoyment, learning, and challenge. In the students' interviews, there was evidence that their enjoyment of the games gave them the confidence to be involved, to compete, to take risks and then to learn, sometimes without realising it.

Furthermore, the teachers noted that students appeared more confident during the repeat lessons when their background number-fact knowledge was sound. This latter observation corresponds with the results of Higgins' (1997) research with similar-aged students, which found that students with good problem-solving knowledge had more confidence in finding a way to start, and in persevering towards a solution. This may have been the case in our research. In our study, the teacher/researcher stressed that it was the thinking process that was important, not the result. With no penalty attached to students' answers, there was no reason to lack confidence in participating in the games and activities physically, socially and cognitively. Garris et al. (2002) argued that this is a real advantage in playing games.

There were also qualitative changes in the teachers' confidence. For example, both teachers acknowledged at the end of the study that they now understood a range of different pedagogical strategies and relevant situations to which 'chance' could be applied. They were more willing and able to extend their students' understanding through appropriate questions, discussion, and de-briefing after playing the games. This study has therefore been successful in providing training, ideas, and support for teachers to teach probability more confidently and in a meaningful way. These were concerns of Watson (2001), Carlson and Doerr (2002), Batenero et al. (2004), and Joliffe (2005) discussed in the introduction to this chapter.

Given the above discussion, confidence has a cognitive association, and is therefore an important aspect in the positive-attitudes–behaviour cycle, when games are used with instruction.

Perceptions of Usefulness On the student survey of attitudes and beliefs, the t-tests found that significantly more students improved their perceptions of the uses of 'chance' (Item 11). The means on Item 3 (useful topic) remained the same. The latter result implies ambivalence towards 'chance' as a topic in mathematics. This aspect may be linked to students' beliefs about mathematics in general. Future exploration of this perspective may be useful. Observation of the extended responses on Items 3 and 11 revealed percentage increases in comments focussing on the relevance of 'chance' (e.g. need it later in life, lots of things involve 'chance', needed for betting and winning). Teachers made similar comments, and noted the need to develop and test strategies, and take risks and wear the consequences. They realised expectations and realities were not always the same. Collectively, the interview students agreed. These results imply that the instruction accompanying the games was necessary and important in highlighting the relevance of probability. This concurs with Jones et al. (1999a, 1999b) in probability research, and Higgins (1997), in problem solving research. Watson (2005) had advocated the importance of context in developing probability understanding.

Anxiety In the student survey of attitudes and beliefs, the t-tests indicated that students felt significantly less anxious at the end of the study (Item 4). There was a non-significant increase on Item 10 ('chance' does not worry me). The extended responses on these items focussed on positive aspects associated with 'chance' being easy and not something to worry about. Teacher A had worried about teaching 'chance', and her ability to take the confusion out of the topic for her students. This stemmed from her own lack of confidence and knowledge, discussed earlier. This anxiety had lessened at the end of the study. Both teachers believed that the nature of the games and activities reduced the anxiety levels for the students, because they (the students) were "engaged from the start". Thus, carefully-selected games and activities have the potential to halt the development of the negative-attitudes–behaviour cycle through anxiety, and instead, to have a positive effect on commitment and involvement (Covington and Omelich 1987).

The teachers did notice anxiety amongst some students, especially the more mathematically weak, when asked to explain their actions and thinking orally or in writing, or when games required a good grasp of number facts. The latter observation was reported by Ritson (1998), who concluded that a weak arithmetic understanding may limit children's probabilistic decision-making ability. Teacher B felt that when students were worried about expressing their views, they were responding to their belief that mathematics related to aiming for the correct answer. While there is no substantive evidence in the study to support this, there is evidence in the literature that beliefs about mathematics influence how students emotionally respond in the classroom (McLeod 1992; Yackel and Rasmussen 2002).

Motivation On the student survey of attitudes and beliefs, the t-tests were significant in finding that students would like to learn more about 'chance' in class (Item 5). There was a non-significant increase in the desire to learn more about 'chance' next year (Item 12). Students' extended responses on these items were positive, focussing on 'chance' being fun and interesting, or aligned with learning (e.g. advantages of learning 'chance', neglected in the curriculum). Generally, the percentages of students giving such responses increased between the pre- and post-tests.

Both teachers were keen to try the new activities with their classes, and stimulated by the approaches used in this study—thinking and reasoning, writing, comparing actual and expected, taking risks, making choices, and valuing students' comments. They thought that the children seemed more committed to their learning, and that the games and activities motivated even the weaker students. At different times, teachers stated that students were keen, competitive, fully engaged, challenged to find better strategies; and the class thinkers spent considerable time writing responses. Interview students echoed many of the teachers' sentiments. In particular, they were motivated by the activity approach as an alternative to "just writing stuff". Sometimes they did not realise they were learning. There was evidence of lower levels of motivation and enjoyment, and less desire to participate without prodding from the teacher, when there was a lack of prerequisite number skills, or there was an element of repetition and monotony in the activity. Mor et al. (2006) noted the link between motivation and participation. Thus motivation is an important element in the positive learning cycle.

Booker (2000) highlighted the value of games in raising levels of motivation to positively influence learning situations. In this study, it was found that the probability games, rather than the probability activities (non-game), were more successful in motivating the students to be involved in their learning. The probability games were obviously novel for the students; whereas the probability activities were probably similar to activities experienced in previous years of school. The contrast may explain the differences in motivation.

4.1 Beliefs About Luck

Lucky People On the student survey of attitudes and beliefs, the t-tests indicated that at the end of the project students had a lower belief in the idea of people being born lucky (Item 6). Fewer students believed in lucky people (Item 14), though this result was non-significant. The majority of extended responses indicated a lack of superstition, and this majority had increased by the post-test. Over 50 percent of responses on Item 14 were based on some understanding of 'chance' (e.g. everyone's luck is the same), yet, despite instruction highlighting discrepancies, some students still depended on their personal observations or experiences. Teacher B had suggested that "lucky" people were those with a positive outlook. At the end of the study, the teachers and some students appreciated that in certain circumstances, good strategies could increase the chance of winning. This lessened the degree of belief in people being lucky.

Lucky Strategies On the student survey of attitudes and beliefs, the t-tests were non-significant for Item 7 (crossing fingers does not improve luck) and Item 13 (improving chances of winning with lucky numbers, birthdays, house numbers). The first result may have been tempered by the negative wording of the question. Again, the majority of extended responses indicated a lack of superstition, and this majority had increased by the post-test. Simultaneously, percentages of responses showing superstition decreased for the major categories. Furthermore, for Item 13, "understanding of 'chance'" (e.g. "if all numbers have an even chance, then lucky numbers don't matter") became a major category on the post-test, with 38 percent of students providing such responses.

From the teacher and student interview data, there was evidence of an appreciation that developing good strategies increased the chance of winning. This implies less dependence on lucky strategies in the future. The beliefs that Truran (1995) found in her study (e.g. belief in holding the dice in a certain way) were less obvious in this study (except in some repeat sessions). Watson et al. (1995), in their study of children's understanding of luck, noted the need for activities "that challenge the concept of luck which implies increased likelihood, and (which) lead students to a concept which implies an appreciation of equal likelihood" (p. 556). The above results show that our study has contributed to the satisfying of this need.

In reviewing a number of studies, Lester (2002) concluded that beliefs are "notoriously resistant to change, even in the face of overwhelming evidence to the contrary" (p. 348). He questioned whether students were responding to the researchers by giving them the responses they wanted, whether the students did actually change their fundamental beliefs, or whether the students learnt that they should have particular beliefs. In the current study, beliefs about lucky people and lucky strategies were challenged through instruction, and also through construction (situations, tables, diagrams). Lester's questions could be valid questions here, and the answers are hard to justify without longitudinal data. The positive outcomes of this research are in the evidence that through instruction, beliefs about luck can become more informed, and linked to a better understanding of 'chance'; and, at the Year 7 level, students are more likely to rationalise from a position that shows they are not superstitious. However, as Fischbein and Gazit (1984) and Konold (1991) found, some students, despite instruction to the contrary, still rationalised from a superstitious perspective.

Despite being of only a short duration, this study has been important in presenting both teachers and students with an alternative way of teaching and learning probability. Both parties appeared to appreciate this. For example, in line with the report on educational games by McFarlane et al. (2002), teachers in this study valued games/activities because they provoked interest and motivation for learning; attention and persistence; and working in a team, discussing, and sharing.

Shaughnessy (1992) and Watson et al. (2004) advocated for the benefits of longitudinal studies, because of the long-term perspective on change. This study makes no declaration regarding the permanency of any changes in attitudes and beliefs, and only reports on the observations made.

The students' extended responses on the survey, the interviews, and teacher observations have been useful in revealing the impact of probability games and activities, implemented in a Year 7 classroom. Collective results show the potential for such games/activities to increase enjoyment, confidence, motivation, and perceptions of usefulness; and lessen anxiety about 'chance' and superstitions about luck. However, it must be remembered that the teacher was not the regular one, the activities were different from usual mathematical activities, the normal classroom routine was interrupted with the presence of extra people, extra time was spent on this topic, and the teaching/learning processes varied from the norm (e.g. thinking, explanation, etc.).

Did cognitive performance improve as a result of the teaching intervention? There is probably no definitive answer without further data. However, there were signs that the games and activities that brought about some changes in attitudes and beliefs also brought some cognitive change. For example, the quantitative results for the students' performance on seven items showed a statistically significant improvement. The teachers thought that the children were involved in their learning, and were thinking about what they were doing, rather than aiming for the correct answer; they were predicting, trying different strategies, learning number facts; many were discussing options, and asking questions. The interview students (i) demonstrated they had acquired strategies to use in 'chance' situations, (ii) were more aware of multiplication number fact combinations, and (iii) stated that at times they did not realise they were learning. Some were challenged to make sense of outcomes through reasoning, experimentation and discussion. The learning process was frustrated when students did not like the game (e.g. boring, repetitive), or perceived it as unfair. Both teachers, especially Teacher A, confirmed they were better equipped with new strategies and better knowledge to teach probability in the future.

Were improvements in attitudes and beliefs associated with improvements in cognitive performance? There was a weak but significant positive correlation between the global attitude score and the total cognitive test score, but again a more definitive answer would be possible with further data. Therefore, the results of this study have implications for future longitudinal studies of this type, and for teaching and learning probability.

Professional Development One important outcome is the realisation that teachers do need assistance in teaching probability, and therefore professional development is recommended. In this study, the teachers did not know what they did not know. Teachers must go beyond the mathematical side of probability, by encouraging reasoning, explaining, experimenting, taking risks, and challenging judgements. This point was taken up by Watson et al. (2004), who advocated the importance of students' descriptive understandings or judgements ("Chance Beliefs") over calculating probabilities ("Chance Measurement"). Such a position requires that teachers have good knowledge of the probability connections to make, the particular aspects of tasks to extend, and the comprehension difficulties that students may face. According to Mor et al. (2006), selecting pedagogically useful games/activities for classroom usage must evolve from a deep understanding of mathematics, games, and pedagogy. Professional development would need to consider all of these aspects.

Students' Prerequisite Knowledge In this study, a lack of prerequisite knowledge, such as poor recall of multiplication facts, acted as a demotivator for students. The required prerequisite skills must be revised in advance to reduce this problem.

Context There are two aspects to be considered with respect to context in this study—the context to which probability ideas are applied, and the classroom environment in which the learning takes place. With respect to the former aspect, the experience and knowledge of the researcher/teacher meant that students had coaching in the applications of probability ideas in everyday life. Higgins (1997) found such instruction necessary. At the end of the study, it was evident that all participants had a better understanding of the many uses and applications of 'chance' ideas. This adds support to the need for professional development to increase knowledge in this area.

With respect to the latter aspect (classroom environment), the important things to consider are the nature of the games and activities selected to support the learning of probability concepts, and the teaching style used to draw out the essential ideas in a non-threatening classroom environment. Contextual factors are important in shaping attitudes and beliefs about mathematics (Greer et al. 2002; Wilkins and Ma 2003; Yackel and Rasmussen 2002; Zan and Martino 2007). In this study, it was the activity approach and the engagement with the games/activities that permitted useful discussions about some of the pitfalls and uses of probability, and the challenge to try different strategies for winning. It was the researcher/teacher's requirement that any answer that allowed thinking, reasoning, and experimenting was okay. The enjoyment factor was prominent. Indeed, an essential characteristic of games is that they are a source of fun and enjoyment (Garris et al. 2002), and they serve as "attractors, to motivate and involve pupils (Mor et al. 2006). Additionally, students' beliefs about luck were challenged with the construction of tables and tree diagrams showing possible outcomes. Lack of prerequisite knowledge appeared to hamper these processes at times. Such contextual factors imply that creating an environment which is enjoyable yet challenging, may, as Wilkins and Ma (2003) state, curb the tendency for negative attitudes to develop, and challenge erroneous beliefs.

The activity approach used in this study shows potential for halting the development of the negative-attitudes–behaviour cycle and encouraging the positive-attitudes–behaviour cycle, and for challenging beliefs relating to learning 'chance'. There are indications that such an approach may lead to improved cognitive performance. Additional data analysis may reveal more definitive results. Future research needs to take an approach where researchers and classroom teachers work together over an extended time period. This paper makes a contribution to the dearth of research on the role of attitudes and beliefs in probability and statistics education (particularly in primary schools), to which Gal et al. (1997) refer. Given the relatively recent emphasis of probability and statistics in the school curriculum, such research is important in understanding the development of attitudes and beliefs through instructional experiences, in order to better inform and plan instructional programs (Gal et al. 1997).

References

Aiken, L. R. (1970). Attitudes toward mathematics. *Review of Educational Research*, *40*(4), 551–596.

Ajzen, I., & Fishbein, M. (2000). Attitudes and the attitude-behavior relation: reasoned and automatic processes. In W. Stroebe & M. Hewson (Eds.), *European review of social psychology* (Vol. 11, pp. 1–33).

Amir, G., & Williams, J. (1999). Cultural influences on children's probabilistic thinking. *The Journal of Mathematical Behavior*, *18*(1), 85–107.

Batenero, C., Godino, J. D., & Roa, R. (2004). Training teachers to teach probability. *Journal of Statistics Education*, *12*(1), 1–19.

Bernstein, W. M., Stephan, W. G., & Davis, M. H. (1979). Explaining attributions for achievement: a path analytic approach. *Journal of Personality and Social Psychology*, *37*(10), 1810–1821.

Bragg, L. (2007). Students' conflicting attitudes towards games as a vehicle for learning mathematics: a methodological dilemma. *Mathematics Education Research Journal*, *19*(1), 29–44.

Bobis, J., Anderson, J., Martin, A., & Way, J. (2011). A model for mathematics instruction to enhance student motivation and engagement. In D. Brahier & W. Speer (Eds.), *Motivation and disposition: pathways to learning mathematics, 73rd yearbook*. Reston: National Council of Teachers of Mathematics (NCTM).

Booker, G. (2000). *The maths game: using instructional games to teach mathematics*. Wellington: New Zealand Council for Educational Research.

Carlson, L., & Doerr, H. (2002). Secondary teachers' attitudes toward probability and the teaching strategies. In B. Barton, K. Irwin, M. Pfannkuch, & M. Thomas (Eds.), *Mathematics education in the South Pacific. Proceedings of the 25th annual conference of the Mathematics Education Research Group of Australasia* (pp. 171–178). Auckland: MERGA.

Cathcart, W. G., Pothier, Y., Vance, J., & Bezuk, N. (2006). *Learning mathematics in elementary and middle schools* (4th ed.). Upper Saddle River: Pearson Education.

Covington, M. V., & Omelich, C. (1987). "I knew it cold before the exam": a test of the anxiety-blockage hypothesis. *Journal of Educational Psychology*, *79*, 393–400.

DeBellis, V. A., & Goldin, G. (2006). Affect and meta-affect in mathematical problem solving: a representational perspective. *Educational Studies in Mathematics*, *63*, 131–147.

Dweck, C. (1986). Motivational processes affecting learning. *The American Psychologist*, *41*(10), 1040–1048.

Estrada (2002). *Analysis of attitudes and elementary statistical knowledge in pre-service teachers*. Unpublished PhD thesis, University of Barcelona.

Fennema, E., & Sherman, J. (1976). *Fennema–Sherman mathematics attitude scales: instruments designed to measure attitudes toward the learning of mathematics*. Washington: American Psychological Association. Psychological Documents (MS No. 1225).

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behaviour: an introduction to theory and research*. Reading: Addison-Wesley.

Fishbein, M., & Gazit, A. (1984). Does the teaching of probability improve probabilistic intuitions? An exploratory research study. *Educational Studies in Mathematics*, *15*, 1–24.

Gal, I. (2005). Towards "probability literacy" for all citizens: building blocks and instructional dilemmas. In G. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 39–63). New York: Springer.

Gal, I., & Ginsburg, L. (1994). The role of beliefs and attitudes in learning statistics: towards an assessment framework. *Journal of Statistics Education*, *2*(2). Available 15 November 2009 from http://www.amstat.org/publications/jse/v2n2/gal.html.

Gal, I., Ginsburg, L., & Schau, C. (1997). Monitoring attitudes and beliefs in statistics education. In I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 37–51). Amsterdam: IOS Press. Chapter 4. Available 15 November 2009 from http://www.stat.auckland.ac.nz/~iase/publications/assessbk/chapter04.pdf.

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: a research and practice model. *Simulation & Gaming*, *33*(4), 441–467.

Greer, B., Verschaffel, L., & DeCorte, E. (2002). "The answer is really 4.5": beliefs about word problems. In G. C. Leder, E. Pehkonen, & G. Törner (Eds.), *Beliefs: a hidden variable in mathematics education?* (pp. 271–292). Dordrecht: Kluwer Academic.

Gresalfi, M. S., & Cobb, P. (2006). Cultivating students' discipline-specific dispositions as a critical goal for pedagogy and equity. *Pedagogies: An International Journal*, *1*(1), 49–58.

Grootenboer, P., & Hemmings, B. (2007). Mathematics performance and the role played by affective and background factors. *Mathematics Education Research Journal*, *19*(3), 3–20.

Higgins, K. M. (1997). The effect of year-long instruction in mathematical problem solving on middle-school students' attitudes, beliefs, and abilities. *The Journal of Experimental Education*, *66*(1), 5–29.

Holmes, E. E. (1990). Motivation: An essential component of mathematics instruction. In T. J. Cooney & C. R. Hirsch (Eds.), *Teaching and learning mathematics in the 1990s, 1990 yearbook* (pp. 101–107). Reston: National Council of Teachers of Mathematics (NCTM).

Joliffe, F. (2005). Assessing probabilistic thinking and reasoning. In G. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 325–344). New York: Springer.

Jones, B., Uribe-Flórez, L., & Wilkins, J. (2011). Motivating students with manipulatives: using self-determination theory to intrinsically motivate students. In D. Brahier & W. Speer (Eds.), *Motivation and disposition: pathways to learning mathematics, 73rd yearbook*. Reston: National Council of Teachers of Mathematics (NCTM).

Jones, G. A., Langrall, C. W., Thornton, C. A., & Mogill, A. T. (1999a). Students' probabilistic thinking in instruction. *Journal for Research in Mathematics Education*, *30*(5), 487–519.

Jones, G., Thornton, C., Langrall, C., & Tarr, J. (1999b). Understanding children's probabilistic reasoning. In L. Stiff & F. Curcio (Eds.), *Developing mathematical reasoning in grades K-12, 1999 yearbook*. Reston: National Council of Teachers of Mathematics (NCTM).

Kamii, C., & Rummelsburg, J. (2008). Arithmetic for first graders lacking number concepts. *Teaching Children Mathematics*, *14*(7), 389–394.

Konold, C. (1991). Understanding students' beliefs about probability. In E. Von Glasersfeld (Ed.), *Radical constructivism in mathematics education* (pp. 139–156). Dordrecht: Kluwer Academic.

Krosnick, J., Judd, C., & Wittenbrink, B. (2005). The measurement of attitudes. In D. Albarracin, B. Johnson, & M. Zanna (Eds.), *The handbook of attitudes* (pp. 21–76). Mahwah: Erlbaum.

Lee, H. S., & Lee, J. T. (2009). Reasoning about probabilistic phenomena: lessons learned and applied in software design. *Technology Innovations in Statistics Education*, *3*(2), 1–21.

Lester, F. K. Jr. (2002). Implications of research on students' beliefs for classroom practice. In G. C. Leder, E. Pehkonen, & G. Törner (Eds.), *Beliefs: a hidden variable in mathematics education?* (pp. 345–353). Dordrecht: Kluwer Academic.

McFarlane, A., Sparrowhawk, A., & Heald, Y. (2002). Report on the educational use of games: an exploration by TEEM [Teachers Evaluating Educational Media] of the contribution which games can make to the education process. *Available*, *25 June 2008* from http://www.teem.org.uk/publications/teem_gamesined_full_pdf.

McLeod, D. (1992). Research on affect in mathematics education: a reconceptualisation. In D. Grouws (Ed.), *Handbook of research on mathematics and learning* (pp. 575–597). New York: Macmillan.

Mor, Y., Winters, N., Cerulli, N., & Björk, S. (2006). Learning patterns for the design and development of mathematical games. Available 19 June 2010 from http://eprints.ioe.ac.uk/4223/1/LP-LitReview-v2.pdf.

Nisbet, S. (2006). Mathematics without attitude. I n *Keynote address to the Annual May conference of the Queensland Association of Mathematics Teachers*, Brisbane (Vol. 6).

Nisbet, S., & Williams, A. (2009). Improving students' attitudes to chance with games and activities. *Australian Mathematics Teacher*, *65*(3), 25–37.

Onwuegbuzie, A. (2000). Statistics anxiety and the role of self-perceptions. *Journal of Educational Research*, *93*(5), 323–330.

Randel, J. M., Morris, B. A., Wetzel, C. D., & Whitehill, B. V. (1992). The effectiveness of games for educational purposes: a review of recent research. *Simulation & Gaming*, *23*(3), 261–276.

Ritson, R. (1998). *The development of primary school children's understanding of probability*. Unpublished thesis, Queen's University, Belfast.

Sánchez, E. S. (2002). Teachers' beliefs about usefulness of simulation with the educational software Fathom for developing probability concepts in statistics classroom. In B. Phillips (Ed.), *Developing a statistically literate society. Proceedings of the sixth international conference on teaching statistics* (ICOTS 6), Cape Town. Voorburg: International Statistical Institute (ISI).

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. A project of the National Council of Teachers of Mathematics*. New York: Macmillan.

Steffe, L. P., & Thompson, P. W. (2000). Teaching experiment methodology: underlying principles and essential elements. In R. Lesh & A. E. Kelly (Eds.), *Research design in mathematics and science education* (Vol. pp, pp. 267–307). Hillsdale: Erlbaum.

Stipek, D., Salmon, J., Givvin, K., Kazemi, E., Saxe, G., & MacGyvers, V. (1998). The value and convergence of practices suggested by motivation research and promoted by mathematics education reformers. In D. Brahier & W. Speer (Eds.), *Journal for Research in Mathematics Education: Vol. 29. Motivation and disposition: pathways to learning mathematics, 73rd yearbook* (pp. 465–488). Reston: National Council of Teachers of Mathematics (NCTM).

Tapia, M., & Marsh, G. (2004). An instrument to measure mathematics attitudes. *Academic Exchange Quarterly*, *8*(2), 16–21.

Truran, K. (1995). Animism: a view of probability behaviour. In B. Atweh & S. Flavel (Eds.), *Galtha. Proceedings of the 18th annual conference of the Mathematics Education Research Group of Australasia* (pp. 537–541). Darwin: MERGA.

Tversky, A., & Kahneman, D. (1973). Availability: a heuristic for judging frequency and probability. *Cognitive Psychology*, *5*, 430–454.

Watson, J. (2001). Profiling teachers' competence and confidence to teach particular mathematics topics: the case of chance and data. *Journal of Mathematics Teacher Education*, *4*, 305–337.

Watson, J. (2005). Is statistical literacy relevant for middle school students? *Vinculum*, *42*(1), 3–10.

Watson, J., Caney, A., & Kelly, B. (2004). Beliefs about chance in the middle years: longitudinal change. In I. Putt, R. Faragher, & M. McLean (Eds.), *Proceedings of the 27th annual conference of the Mathematics Education Research Group of Australasia. Mathematics education for the third millennium: towards 2010* (pp. 581–588). Sydney: MERGA.

Watson, J., Collis, K. F., & Moritz, J. (1995). Children's understanding of luck. In B. Atweh & S. Flavel (Eds.), *Proceedings of the 18th annual conference of the Mathematics Education Research Group of Australasia. Galtha* (pp. 550–556). Sydney: MERGA.

Watson, J., Collis, K. F., & Moritz, J. (1997). The development of chance measurement. *Mathematics Education Research Journal*, *9*(1), 60–82.

Wilkins, J. L. M., & Ma, X. (2003). Modelling change in student attitude toward and beliefs about mathematics. *The Journal of Educational Research*, *97*(1), 52–63.

Williams, A. M. (1992). *An investigation into the relationship between attitudes and performance in an introductory tertiary statistics subject*. Masters thesis, Queensland University of Technology, Brisbane.

Wood, M. (2005). The role of simulation approaches in statistics. *Journal of Statistics Education*, *13*(3), 1–10.

Woolfolk, A., & Margetts, K. (2007). *Educational psychology*. French's Forest: Pearson Education.

Yackel, E., & Rasmussen, C. (2002). Beliefs and norms in the mathematics classroom. In G. C. Leder, E. Pehkonen, & G. Törner (Eds.), *Beliefs: a hidden variable in mathematics education?* (p. 313). Dordrecht: Kluwer Academic.

Zan, R., & Di Martino, P. (2007). Attitudes towards mathematics: Overcoming positive/negative dichotomy. In *The Montana Mathematics Enthusiasts Monograph* (Vol. 3, pp. 157–168). Montana: The Montana Council of Teacher of Mathematics.

Section IV Commentary: The Perspective of Mathematics Education

Jane M. Watson

The ten chapters selected for this section were chosen to reflect the perspective of Mathematics Education. The first task of this commentary is hence to reflect on this perspective in comparison or contrast with the perspectives of the other three sections of the book: Mathematics and Philosophy, Psychology, Stochastics. The next consideration is how the topics of the ten chapters fit within the perspective of Mathematics Education and the contributions they make to our understanding of probability within Mathematics Education. This leads to further suggestions for extension of the projects and issues covered in the chapters. Finally, some comments are made about other past, current, and potential contributions from researchers in probability to the field of Mathematics Education.

1 Why Mathematics Education?

In line with the title of the series of which this volume is a part, "*Advances in Mathematics Education*," on the one hand, it seems appropriate to label the final section, Mathematics Education. On the other hand, the question then arises in relation to how the contributions in this section are different from those in the other three sections if the entire series is about Mathematics Education.

To think about this question, it is necessary to explore the scope of Mathematics Education itself. Unfortunately, in this series it appears to be taken as an undefined term. For the purposes of this commentary, Mathematics Education is assumed to encompass broadly the "teaching and learning of mathematics." Although teaching usually comes first in such a phrase, it is learning that is the goal of Mathematics Education. Students and teachers are hence the focus of studies in Mathematics Education. The environments in which they interact, however, influence the outcomes of

J.M. Watson (✉)
University of Tasmania, Hobart, Tasmania, Australia
e-mail: Jane.Watson@utas.edu.au

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_37,

the education process. These environments include the local culture, language and learning issues, and economic conditions, as well as the all-important curriculum set down by education authorities.

As well as the environments within which the learning and teaching take place, a significant factor is the method by which the mathematics, in this case probability, becomes a part of the learner's repertoire. Over time suggested methods have evolved, reflecting to some extent the impact of research, for example, into the detail of students' development of conceptual understanding and into the requirements of the pedagogy of teachers that will assist that development. These broad interpretations of Mathematics Education allow the suggestion that the initial section of this book, on Mathematics and Philosophy, contributes to Mathematics Education in relation to curriculum content and perhaps pedagogy. The section on Psychology definitely contributes to the teacher's appreciation of the learner's development of understanding, at times supplemented by suggestions for the curriculum. The section on Stochastics, in developing links with statistics, not only provides further implications for the curriculum, but also offers advice on teaching, including the use of technology.

If these other sections contribute to the position of probability within Mathematics Education, how does the final section contribute something extra? It appears that the chapters in this section were mainly chosen because they focus more specifically on some feature of learning probability that is classroom based and the product of the author/s' research. Except for the chapter on frameworks by Mooney, Langrall and Hertel, all of the chapters in this section present data of some type related to interventions with groups of learners ranging from age $5\frac{1}{2}$ years to practicing secondary teachers. Four chapters deal with elementary school learners, three report on high school students, one discusses longitudinal development from Grade 4 to graduate school, and one considers the thinking of secondary teachers. The topics range from basic concepts such as sample spaces and equally likely outcomes to the law of large numbers, with problems related to binomial and conditional probabilities considered at the high school level.

2 Contributions to This Volume

In grouping the chapters in the section on Mathematics Education, four perspectives from the probability curriculum are addressed: attitudes and beliefs, with the potential for a cultural influence on them; the necessity to understand and appreciate the sample space as underlying the calculation of probabilities; the law of large numbers; and two advanced secondary topics, binomial and conditional problems. In various ways these chapters, as well as the theoretical chapter on frameworks, are a representative sample of the classroom research carried out in the past decade or two. Only the chapter by Mamolo and Zazkis begins to move beyond this focus into considering the educational imperative to look beyond students and content to the realms of teachers and the pedagogical content knowledge required to address the needs of learners as they are developing the understanding exemplified in the other chapters.

What do we learn and what further questions should be asked at the conclusion of the studies reported here? In terms of attitudes and beliefs towards probability and how they are influenced by several chance games or activities, Nisbet and Williams provide much detail but it would be very useful in assessing outcomes if further information were also made available, for example, about what actually happened in the intervention in terms of instructions given or worksheets provided to students, why the exact same lesson was presented twice (by the researcher and then the classroom teacher) and how that affected the interest level of the students, and what probability content was expected to be learned by students. Although "test" questions were asked, there is no indication of their content and their relationship to the games played. The issue of playing games and gaining insight into probabilistic understanding has been considered for a long time (e.g., Bailey 1981; Bright et al. 1981) and perhaps a longer intervention that could document transfer, for example, like that carried out by Prediger and Schnell, would lead to even more positive outcomes for attitudes and beliefs.

Sharma presents a different perspective on beliefs about probability in looking at their direct influence on responses to probability questions rather than their influence on general views of chance. Beliefs about events derived from Fijian culture are used to explain students' interview responses to questions about equally likely outcomes and independence. Because studies from other cultures have found similar results (e.g., Watson et al. 1997), the question arises about the relative frequency of the types of responses in different cultural settings around the world. Although suggestions are made for teacher interventions, the possibility of knowing where the students' cultural beliefs begin may not be all that helpful in moving them meaningfully into the abstract mathematical sphere. There is the danger of students learning that there is one set of rules in the world of mathematics and another set when they go home to discuss chance events with their parents. This is reminiscent of the pre-service teacher known to the author who explained completely correctly the theory behind tossing a coin repeatedly but added, "But I will always call 'tails' because it is lucky for me." Throughout this volume pleas are made for a curriculum that specifically addresses three perspectives on probability—theoretical, frequency, and subjective. Listed in this order there is implicit support for a hierarchical order of importance with theoretical at the top and subjective at the bottom. Sharma and other authors in this section raise the importance of this triad and lament the lack of research interest, Sharma in particular, in the subjective sphere. In thinking from the students' perspectives of beginning as subjective learners, rather than from the historical perspective of probability as pure mathematics to be embedded in students' minds, perhaps the teachers' knowledge of "student as learner" can contribute to an improved pedagogical content knowledge in transitioning from subjective, to frequentist, to theoretical probability understanding.

Once moving past the attitudes and beliefs that can influence outcomes related to determining probabilities, what is the foundation concept without which Mathematics Education cannot hope to build meaningful student understanding of probability?

Excluding the chapters on technical problems (binomial and conditional) it is possible to imagine a fierce debate between the authors of the three chapters focusing on sample spaces and the authors of the two chapters on the law of large numbers. It is not possible to calculate probabilities without valid sample spaces but the point of practical probability is lost without the law of large numbers.

The importance of sample space to mathematics education as seen through the eyes of researchers is highlighted in the varied approaches in the three chapters on the topic in this section. Nilsson "begins at the very beginning," describing the authentic environment of a Grade 6 classroom where students are asked to determine the sample space, and hence the chance of picking a yellow piece of candy from a bag, by repeated replacement sampling. Hence the students start within a frequency/experimental context to predict the theoretical context, which can ultimately be determined by opening the bag to count the contents. For Nilsson the sample space is the link between sample and population and although increasing sample size is a factor, the law of large numbers is not an explicit feature. The proportion of yellow pieces within the sample size used (number of single draws with replacement) becomes the salient issue for the Grade 6 students. Although the conceptual expectation is sophisticated, the context is basic and does not change throughout the teaching. A follow-up research question could usefully address the transfer of learning to another context.

In contrast to the Nilsson approach, Maher and Ahluwalia present a longitudinal cross-sectional study that focuses much more specifically on determining the specific elements of a sample space only from a theoretical perspective based on counting techniques and combinatorial reasoning. The increased sophistication of arguments is mathematical rather than statistical in nature; hence, model building rather than sampling is the mechanism for reaching results. Starting when the students are in Grade 4, they are allowed to explore, interact, and create models, which are then rejected, accepted, or extended by their fellow students over the years until the last description is of one of the participants as a graduate student. In contrast to Nilsson who uses only one context for the consideration of the sample space, Maher and Ahluwalia use several settings to follow increasingly complex configurations to model the sample spaces. In beginning and ending in the world of theoretical probability, the issue of the law of large numbers does not arise.

Although still based on a sample space, the chapter by Mamolo and Zazkis provides a very different perspective in setting up an infinite sample space of all real numbers between 0 and 10. Teachers are then given the task to adjudicate an imagined argument between two students about the probably of one of them identifying the secret number chosen by the other. Although any mathematical resolution is highly theoretical and based in understanding the infinite cardinality of the real numbers, the discussion contextualizes the responses in terms of the teachers' conceptualizations of randomness, the meaning of "real" in real numbers, the concreteness or otherwise of real numbers, and the relationship of a "zero" chance and impossibility. In considering the various potential beliefs of students about the context presented, the message becomes similar to that of Sharma in relation to the Fijian culture that also influences students' decisions on answers to probability problems. In the case of Mamolo and Zazkis, the culture rests in the level of understanding of the mathematics of infinity and infinite sets, and it must be assumed that appreciating sets of measure zero is likely to be beyond the reach of school students and most of their teachers. As a topic such as this is not in any school curriculum, the practicalities of dealing with it might be questioned within the usual realms of Mathematics Education. The apparent paradox of dealing with actual numbers such as 3 or 6.79 and saying they have a probability of zero of being guessed by another person, is similar to the paradox of the expected number of children in families if they continue to have children until the family has an equal number of boys and girls. Any family that is imagined will eventually be of finite size but the expected (average) number of children is infinite (Straub 2010). The culture of social understanding in Fiji and the culture of pure mathematical understanding in the real number system are very different but both can be seen as significant influences on decision making in probabilistic situations.

The approaches to explore children's understanding of the law of large numbers by Pararistodemou and by Prediger and Schnell both are based on computer game simulations, which facilitate completing a large number of trials more quickly than doing them by hand. More details of the goal and operation of the space game and how situated abstractions are developed by the children would have been helpful in interpreting the outcomes related to students' expressed robust intuitions for the law of large numbers, as presented by Pararistodemou. The students' responses in the Prediger and Schnell study are analyzed with a detailed scheme that follows the development of their appreciation of the importance of a large number of trials through their explanation of bets on which of four "animals" will win a race. The distinction between the internal number of trials within a game and the total number of games played is useful in tracking development because each can be separated (large and small) to create a two-dimensional grid. The question might be asked of the ultimate importance of the large number of trials over a large number of games but the analysis suggests the students who were the focus of the investigation realized this. It would be interesting to apply this model in an analysis of the Pararistodemou data as one suspects it would support the conclusions claimed there. The ultimate question for both of these studies is: What is the evidence of transfer of understanding of the law of large numbers to other settings? This question should provide a starting point for future research.

The chapters of Sanchez and Landin and of Huerta focus on particular, more advanced topics in the high school probability curriculum, namely binomial problems and conditional probability problems. Sanchez and Landin focus on a structural approach to analyzing students' responses to the problems in a manner similar to that suggested by Mooney et al. at the beginning of the section. Although structures are designed to display increasingly complex understanding, if the highest level of response is the "correct" answer that may be shown in a strictly procedural fashion without any accompanying description of the associated understanding, there may be a question of whether the solution should genuinely be regarded as the most complex structurally. Although the potential confounding by the alternate statement of the question is noted in explaining the results, this is not clarified to the extent of distinguishing between two possibilities in detail. A further discussion of the implications of the results, either in terms of teaching or in terms of desirable problem statements, would be useful to teachers.

In a similar fashion, Huerta presents an even more detailed account than Sanchez and Landin, of the solving of conditional probability problems. Using a trinomial graph to map the relationships among the components of a ternary problem, he presents an exceedingly complex model. Again the model is used to classify student solutions to conditional problems and particular areas of error are highlighted. Although interesting from a theoretical point of view, Huerta gives no suggested teaching intervention or trajectory to avoid the difficulties observed and claims the problems arise from the "usual teaching" based on techniques. This is unfortunate given that the claim is also made that "the most efficient [approach] was the arithmetical one of using a 2×2 table." Both Huerta and Sanchez and Landin are concerned with obtaining the correct answers to conventionally stated problems, and although this is obviously important, a display of a correspondingly high level of descriptive understanding is also necessary.

As noted in relation to several of the chapters, when employing concrete contexts such as games (either hands on or with technology) to introduce probability concepts, it is necessary to conduct follow-up research to find out if the understanding is meaningful and transferrable. The dilemmas are sometimes explored within the original contexts (e.g., Prediger and Schnell) and Maher and Ahluwalia provide both this kind of evidence and evidence of transfer across contexts in their longitudinal study. Transfer is a continuing issue with research into both students' learning and teachers' pedagogical content knowledge.

It is also possible to ask whether some of the chapters from other sections of this volume can be claimed to contribute to Mathematics Education in the sense of improving learning for students. The chapter by Jolafee, Zazkis, and Sinclair, for example, reports on learners' descriptions of randomness and the chapter by Martignon suggests the introduction of tools for dealing with risk and decision-making under uncertainty. Careful reading of other chapters is also likely to reveal hints for improved classroom practice.

Mooney, Langrall, and Hertel provide a transition of thinking, from the section focusing on Psychology to this one on Mathematics Education. They claim that in his observations of children Piaget was not interested in the influence of education on the development of the thinking he observed and note the influence of Fischbein in moving to the need to acknowledge the interaction of developmental thinking with the context of education. Biggs and Collis (1982), for example, extended Piaget's idea to develop a "structure of observed learning outcomes" (SOLO) without necessarily going behind the observations to make assumptions about their origin. Such a developmental framework, if understood by teachers can be used to observe a current level of development and devise learning activities to assist students in moving to higher levels. From a Mathematics Education perspective, this is the usefulness of such frameworks that have their basis in Psychology. In using such a framework to describe solutions to binomial problems, Sanchez and Landin could provide a useful extension to their research by suggesting detailed specific classroom interventions to assist in moving responses to higher levels. Further, the research of Huerta on conditional probably could be more useful for teachers in planning lessons if he included a similar developmental framework with details of students' increasingly appropriate steps toward solutions. It may be that other authors, such as Martignon and Chilsi and Primi, who write in the Psychology section, provide contributions in this area.

At the end of their chapter, Mooney et al. discuss the influence of probability frameworks on the recently released Common Core State Standards for Mathematics (CCSSM) (Common Core State Standards Initiative 2010) in the United States. It would be useful to consider similar initiatives in other countries. In New Zealand, for example, the title of the country's Mathematics curriculum was changed in 2007 to Mathematics and Statistics, reflecting the importance of statistics for students entering the world of the twenty-first century (Ministry of Education 2007). At every level of the curriculum, one to eight, Probability is given a similar heading to Statistics under the overall Statistics section. Although not specifically based on a particular framework, the content reflects research into probabilistic understanding over the last few decades. Similarly *The Australian Curriculum: Mathematics* (Australian Curriculum, Assessment and Reporting Authority 2013) for Years Foundation to 10 contains one of three major sections on Statistics and Probability, with Chance as a subheading for content at every year level. Although not as research-based as the New Zealand document, it recognizes the development of intuitive ideas in the early years, which is not acknowledged in the CCSSM content. The approach of the CCSSM to Probability appears more mathematical (i.e., theoretical) than would seem appropriate from the research findings of recent years, particularly those that illustrate the development of understanding of variation and expectation in quite young children (e.g., Watson 2005).

In a recent review of children's understanding of probability that could also contribute to curriculum development and classroom planning, Bryant and Nunes (2012) cover the four cognitive demands they believe are made on children learning about probability: randomness, the sample space, comparing and quantifying probabilities, and correlations. Although somewhat limited in scope and sources, the review highlights some of the issues that are noted elsewhere in this commentary and require further research. There is a need for more detailed analysis of classroom interventions (e.g., Nilsson; Prediger and Schnell) and of longitudinal studies (e.g., Maher and Ahluwalia). Again this reflects the call of Mooney et al. for considering the practical side of applying the psychological frameworks they introduce. This approach is indeed the Mathematics Education perspective on probability that contributes to this volume in the "*Advances in Mathematics Education*" series.

3 Other Aspects and Future Directions

What are some of the other aspects of Probability education within the scope of Mathematics Education that deserve attention of researchers and educators? Although throughout most of the twentieth century, probability was treated as pure mathematics in school curricula, the necessity for probability in making decisions about the confidence one has in carrying out statistical tests, has meant that these ideas, often treated informally, have filtered down to the school curriculum. Informal probability has had to be applied to decision making. This intuitive conceptual use of probability in judging uncertainty is not related to counting elements in sample spaces or calculating basic probabilities but is equally important. It may involve frequency representations or even subjective decision making with limited evidence but it provides a foundation for a critical use of probability in formal inference later. Makar and Rubin (2009) provide an excellent framework for informal inference to get students started: *generalizations* for populations are made based on *evidence* collected from samples, acknowledging the *uncertainty* in the claim. Studies around students' informal decision making and uncertainty are beginning to appear on the research scene (e.g., Ben-Zvi et al. 2012).

Again in looking behind the formal calculation of probabilities of events, there is the issue of the relationship of expectation and variation as students are developing their conceptions. As noted by Prediger and Schnell, students experience variation in their small sample trials, which at times confounds their predictions of outcomes. Given that the traditional exposure in the school curriculum has been to expectation as probability (or averages) first and to variation as standard deviation much later, the research of Watson (2005) with students from ages 6 to 15 years suggests, perhaps surprisingly, that students develop an intuitive appreciation of variation before an intuitive appreciation of expectation. The complex development of interaction between the two concepts has been shown based on in depth interviews with students employing protocols in chance contexts, as well other contexts relevant in statistics (Watson et al. 2007). This fundamental conflict of variation and expectation underlies decision making and is as important a topic to explore as is the law of large numbers, with which there are connections.

Except for the chapter by Mamolo and Zazkis, which looks at the content knowledge of secondary teachers, the chapters in this section focus on school student learning, its development, associated success, attitudes and beliefs. Since the seminal work of Green (1983), students' understandings and misunderstandings of chance, probability and randomness have been a focus of research. This and the suggestion of developmental models (e.g., Mooney et al.) have contributed much to curricula around the world and to suggestions for teaching, for example, from the National Council of Teachers of Mathematics in the US (e.g., Burrill 2006). In other volumes (e.g., Jones 2005), suggestions are made explicitly as to how teachers can assist learners in developing their probabilistic understanding. Pratt (2005), for example, provides ample justification for the use of technology in terms of building models, testing conjectures, completing large-scale experiments, and considering various contexts linked to basic models. Other researchers make specific suggestions for building concepts based on observed developmental frameworks (such as those reviewed by Mooney et al.) to cover topics such as compound events, conditional probability and independence (Polaki 2005; Tarr and Lannin 2005; Watson and Kelly 2007, 2009). Although a step in the right direction, in contributing to teachers' knowledge of students as learners and the difficulties they are likely to meet, there is still the dilemma of the actual teacher encounter with the confused or blank look of a student who is stuck or has no idea where to start. Prediger and Schnell begin to offer ideas in this regard with their detailed case study of two learners with a betting game.

Compared to student understanding of probability, much less is known about teachers' development of understanding for teaching, especially recognizing how broad it must be in the light of the work of Shulman (1987) on the seven types of knowledge required for teaching and the adaptations by Ball and colleagues for "Mathematical knowledge for teaching" (e.g., Hill et al. 2004). Many variations on these beginnings have been suggested and in some instances the pedagogical content knowledge (PCK) of Shulman has been expanded to encompass not only "pedagogical" knowledge and "content" knowledge and their interaction, but also knowledge of students as learners (Callingham and Watson 2011). Given the previous decades of research on students' errors and development of understanding, knowledge of students as learners must be a critical component of the knowledge teachers bring to the classroom. One approach to exploring such knowledge in probability is based on either asking teachers to suggest both appropriate and inappropriate responses to particular questions or asking teachers to suggest remedial action for authentic inappropriate student responses to questions (Watson and Nathan 2010). This approach may help teachers become aware of the potential difficulties students experience but there are also the issues of planning for teaching episodes and responding on-the-fly to students' unusual classroom contributions. These aspects cannot be explored in teacher interviews or surveys, but must be observed in real time in classrooms. Brousseau et al. (2002) presented an early detailed account describing interactions when teaching probability and statistics in Grade 4, without specific reference to PCK but with insights into what students needed in order to take on the desired understanding. Similarly, observing two Grade 5 classrooms for two lessons on probability, Chick and Baker (2005) found the teachers differed in their content knowledge as well as their PCK and, although there were interactions with students noted, these exchanges were not documented in detail with the knowledge of students as learners being made explicit. It is the ability to react meaningfully on-the-spot in response to a student's answer or question that combines content knowledge and pedagogical knowledge with knowledge of students as learners. Being well-read on the research into students' potential errors and developmental progressions is a great help to teachers but much experience is needed in starting "where the student is at" and not making assumptions based on a formulaic description from a text book.

More research is needed on all aspects of teachers' PCK with respect to probability including documented experiences with students in actual classrooms. This is likely to be time-consuming with much non-relevant material included in recordings. Perhaps a starting point could be based on earlier research related to cognitive conflict (Watson 2002, 2007), where students are presented with genuine conflicting responses from other students and asked to respond. Such genuine student responses could be used as starting points with pre-service or in-service teachers in workshops where they debate the best way to respond. A model for this is provided by Chernoff and Zazkis (2011) with pre-service teachers. When given an inappropriate response to a sample space problem the pre-service teachers could only offer didactic suggestions that did not appreciate the starting point for the student response. It was not until the pre-service teachers were themselves given a sample space problem they could not solve and were dissatisfied with a similar response from the lecturer that they saw the point of "starting where the learner is at." It seems likely that the chapter by Jolafee, Zazkis, and Sinclair earlier in this volume may provide further examples in relation to children's ways of talking about randomness.

Mathematics Education is about teaching and learning. It would appear that finding evidence of facets of student learning has been the major focus of research based around classrooms. Deep thinking needs to occur into the place of teaching in mathematics education and how it can be enhanced to provide improved support for the learning of probability.

4 Final Comment

The existence of a volume of this size solely devoted to Probability in a series on *Advances in Mathematics Education* is a measure of the increased interest in research on the understanding, learning, and teaching of probability in the last 30 years. This growth reflects a growing academic field of Mathematics Education, as well as an intrinsic interest of researchers in the complex topic of probability. The appearance of probability, with one interpretation or another, in the curricula of many countries for over 25 years (e.g., NCTM 1989), is evidence of a synergistic relationship of these two perspectives of Mathematics Education: the outcomes of research have influenced curriculum writers of the importance of probability for school students and its presence in the curriculum has encouraged more research on the topic. It is critical that this continued research focus on teachers as well as students with the aim of producing a generation of citizens who understand the chances and risks involved in the decisions they make.

References

Australian Curriculum, Assessment and Reporting Authority (ACARA) (2013). *The Australian curriculum: mathematics, Version 4.1, 1 February 2013*. Sydney: ACARA.

Bailey, E. (1981). Triples: a game to introduce the facts of chance. In A. P. Shulte (Ed.), *Teaching statistics and probability. 1981 yearbook* (pp. 64–69). Reston: National Council of Teachers of Mathematics.

Ben-Zvi, D., Aridor, K., Makar, K., & Bakker, A. (2012). Students' emergent articulations of uncertainty while making informal statistical inferences. *ZDM Mathematics Education*, *44*, 913–925.

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: the SOLO taxonomy*. New York: Academic Press.

Bright, G. W., Harvey, J. G., & Wheeler, M. M. (1981). Fair games, unfair games. In A. P. Shulte (Ed.), *Teaching statistics and probability. 1981 yearbook* (pp. 49–59). Reston: National Council of Teachers of Mathematics.

Brousseau, G., Brousseau, N., & Warfield, V. (2002). An experiment on the teaching of statistics and probability. *Journal of Mathematical Behavior*, *20*, 363–411.
Bryant, P., & Nunes, T. (2012). *Children's understanding of probability: a literature review*. London: Nuffield Foundation.
Burrill, G. F. (Ed.) (2006). *Thinking and reasoning with data and chance. Sixty-eighth yearbook*. Reston: National Council of Teachers of Mathematics.
Callingham, R., & Watson, J. (2011). Measuring levels of statistical pedagogical content knowledge. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics—challenges for teaching and teacher education: a joint ICMI/IASE study* (pp. 283–293). Dordrecht: Springer.
Chernoff, E. J., & Zazkis, R. (2011). From personal to conventional probabilities: from sample set to sample space. *Educational Studies in Mathematics*, *77*, 15–33.
Chick, H. L., & Baker, M. (2005). Teaching elementary probability: not leaving it to chance. In P. C. Clarkson, A. Downton, D. Gronn, M. Horne, A. McDonough, R. Pierce, & A. Roche (Eds.), *Building connections: theory, research and practice. Proceedings of the 28th annual conference of the Mathematics Education Research Group of Australasia* (pp. 233–240). Sydney: MERGA.
Common Core State Standards Initiative (CCSSI) (2010). Common Core State Standards for Mathematics. Washington, DC: National Governors Association for Best Practices and the Council of Chief State School Officers. Available at: http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf.
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.), *Proceedings of the first international conference on teaching statistics* (Vol. 2, pp. 766–783). Sheffield: Teaching Statistics Trust.
Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics for teaching. *Elementary School Journal*, *105*, 11–30.
Jones, G. A. (Ed.) (2005). *Exploring probability in school: challenges for teaching and learning*. New York: Springer.
Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, *8*(1), 82–105.
Ministry of Education (2007). *Draft mathematics and statistics curriculum*. Wellington: Author.
National Council of Teachers of Mathematics (1989). *Curriculum and evaluation standards for school mathematics*. Reston: Author.
Polaki, M. V. (2005). Dealing with compound events. In G. A. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 191–214). New York: Springer.
Pratt, D. (2005). How do teachers foster students' understanding of probability? In G. A. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 171–189). New York: Springer.
Shulman, L. S. (1987). Knowledge and teaching: foundations of the new reform. *Harvard Educational Review*, *57*, 1–22.
Straub, A. (2010, October 25). How big are families with the same number of girls and boys? Available at: http://arminstraub.com/mathematics/how-big-are-families-with-the-same-number-of-girls-and-boys.
Tarr, J. E., & Lannin, J. K. (2005). How can teachers build notions of conditional probability and independence? In G. A. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 215–238). New York: Springer.
Watson, J. M. (2002). Inferential reasoning and the influence of cognitive conflict. *Educational Studies in Mathematics*, *51*, 225–256.
Watson, J. M. (2005). Variation and expectation as foundations for the chance and data curriculum. In P. Clarkson, A. Downton, D. Gronn, M. Horne, A. McDonough, R. Pierce, & A. Roche (Eds.), *Building connections: theory, research and practice. Proceedings of the 28th annual conference of the Mathematics Education Research Group of Australasia* (pp. 35–42). Sydney: MERGA.
Watson, J. M. (2007). The role of cognitive conflict in developing students' understanding of average. *Educational Studies in Mathematics*, *65*, 21–47.

Watson, J. M., & Kelly, B. A. (2007). The development of conditional probability reasoning. *International Journal of Mathematical Education in Science and Technology*, *38*, 213–235.

Watson, J. M., & Kelly, B. A. (2009). Development of student understanding of outcomes involving two or more dice. *International Journal of Science and Mathematics Education*, *7*, 25–54.

Watson, J., & Nathan, E. (2010). Assessing the interpretation of two-way tables as part of statistical literacy. In C. Reading (Ed.), *Data and context in statistics education: towards an evidence-based society. Proceedings of the 8th international conference on the teaching of statistics*, Ljubljana, Slovenia, July. Voorburg: International Statistical Institute. Refereed paper. Available at: http://www.stat.auckland.ac.nz/~iase/publications/icots8/ICOTS8_5E1_WATSON.pdf.

Watson, J. M., Collis, K. F., & Moritz, J. B. (1997). The development of chance measurement. *Mathematics Education Research Journal*, *9*, 60–82.

Watson, J. M., Callingham, R. A., & Kelly, B. A. (2007). Students' appreciation of expectation and variation as a foundation for statistical understanding. *Mathematical Thinking and Learning*, *9*, 83–130.

Commentary on *Probabilistic Thinking: Presenting Plural Perspectives*

Egan J. Chernoff and Bharath Sriraman

> Martin Gardner's writing is amazingly accurate and reliable. The fact that he made a mistake is simply a testimonial to the difficulty of the [Two Child] problem.
>
> (Khovanova 2011, p. 1)

> But, to my surprise, Erdös said, "No, that is impossible, it should make no difference"... Erdös objected that he still did not understand the reason why, but [after being shown a simulation of the Monty Hall Problem] was reluctantly convinced that I was right.
>
> (Vazsonyi 1999, p. 18)

The above quotations and, more notably, the individuals involved, help cement the popular notion that probability is counterintuitive—just "Ask Marilyn". However, as demonstrated throughout this volume, counterintuitiveness is but one of many different characteristics of probabilistic thinking.

Those of you familiar with research investigating probabilistic thinking in the field of mathematics education, might, at this point in the book, be expecting a "wish list" for future research, which has become customary (e.g., Kapadia and Borovcnik 1991; Jones et al. 2007; Shaughnessy 1992); however, we will not be adding to the list of wish lists. Instead, we have decided to, in this commentary, highlight some of the overarching themes that have emerged from the significant amount of research housed in this volume. Themes emerging from each of the four main perspectives—Mathematics and Philosophy, Psychology, Stochastics and Mathematics Education—are now commented on in turn.

E.J. Chernoff (✉)
University of Saskatchewan, Saskatoon, Saskatchewan, Canada
e-mail: egan.chernoff@usask.ca

B. Sriraman
The University of Montana, Missoula, MT, USA
e-mail: SriramanB@mso.umt.edu

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_38,

1 Perspective I: Mathematics and Philosophy

There are three themes emerging from Perspective I: Mathematics and Philosophy that we wish to highlight. The themes are: different philosophical interpretations of probability, modeling, and subjective probability.

The classical, frequentist, and subjective interpretations of probability, which we denote *The Big Three*, are addressed (to varying degrees) in nearly every chapter and in every perspective of this volume. Worthy of note, and not something that we would consider censoring with our editorial hand, this volume continues to see different researchers utilizing different terminology when referring to The Big Three (see, for example, Borovcnik and Kapadia, this volume). Although not censored, issues inherent in the use of different terminology are discussed in this volume (see, for example, Chernoff and Russell). While the terminology associated with The Big Three has not been solidified in this volume, many authors, within their chapters, have clearly defined their philosophical positions relative to The Big Three. Other authors in this volume, (perhaps) heeding the repeated calls for unification found in the mathematics education literature, utilized (all of) The Big Three interpretations of probability in their research (e.g., Borovcnik and Kapadia; Eichler and Vogel; Pfannkuch and Ziedins). By embracing multiple interpretations of probability, this research further reveals the next major theme that emerged from the first perspective.

Beyond, but related to The Big Three (philosophical interpretations of probability), a second theme that emerged from the first perspective was modeling. As stated by Borovcnik and Kapadia (this volume): "Philosophical difficulties have been prevalent in probability since its inception, especially since the idea requires modeling—probability is not an inherent property of an event, but based on the underlying model chosen" (p. 7). Modeling was found to varying degrees in numerous chapters (e.g., Batanero et al.): some with a heavier emphasis on technology and simulation than others (e.g., Eichler and Vogel; Pfannkuch and Ziedins, Prodromou; Lesh); and some making finer distinctions between reality (Borovcnik and Kapadia; Pfannkuch and Ziedins), virtual reality (Eichler and Vogel), or some sanitized version of reality (Greer; Sriraman and Lee). Although the unification of The Big Three may be achieved through modeling, issues associated with subjective probability still persist: the final theme we will comment on from Perspective I.

As evidenced throughout this volume, and unlike the classical and frequentist interpretations of probability, research in mathematics education has yet to adequately define the subjective interpretation of probability. However, Sharma (this volume) has put forth a call, echoed by Watson (this volume), for research investigating subjective probability. We further echo this call, but do so under the caveats presented by Chernoff and Russell (this volume). Addressing the issues surrounding subjective probability may help the field to move beyond The Big Three resulting in new research embracing "new" interpretations of probability (e.g., propensity—mentioned by Borovcnik and Kapadia, this volume). Like Perspective I, different interpretations of probability also underpin themes emerging from Perspective II.

2 Perspective II: Psychology

Numerous themes emerge from the second perspective of this volume. Of these themes, we have selected two that will now be commented on in turn: the research of Daniel Kahneman and Amos Tversky (and colleagues); and the research of Gerd Gigerenzer (and colleagues).

The original heuristics and biases program of Daniel Kahneman and Amos Tversky (e.g., Kahneman et al. 1982) is seminal to those investigating probabilistic thinking in the field of mathematics education. However, Chernoff (2012), while acknowledging exceptions (e.g., Leron and Hazzan 2006, 2009; Tzur 2011), argues that the mathematics education community has largely ignored more recent developments from the field of cognitive psychology. As a result, developments associated with the original heuristics and biases program (e.g., Gilovich et al. 2002; Kahneman 2011) are not found in mathematics education literature investigating probabilistic thinking. In particular, Chernoff highlights an "arrested development of the representativeness heuristic" (p. 952) in the field of mathematics education, which, if not for certain chapters in this volume (Chiesi and Primi; Ejersbo and Leron; Savard) and the previously mentioned exceptions, may have been extended to heuristics, in general, in mathematics education. Other chapters in this perspective, considering a different notion of heuristics (e.g., Meder and Gigerenzer, this volume), further thwart continuation of this arrested development of heuristics.

The question of whether man is an intuitive statistician, central to the research of Gerd Gigerenzer and colleagues (and, for that matter, Kahneman and Tversky), is found in numerous chapters throughout this volume (e.g., Abrahamson; Brase, Martinie and Castillo-Grouw; Martignon; Meder and Gigerenzer; Saldanha and Liu; Van Dooren). Prior to this volume, the research of Gerd Gigerenzer and colleagues has (inexplicably) been largely ignored by the mathematics education community. We say "inexplicably" because of the two main topics central to the research of Gigerenzer and colleagues: heuristics and risk. Research found in this volume (e.g., Meder and Gigerenzer; Martignon), potentially, signals the dawn of a new era of research for those investigating probabilistic thinking in the field of mathematics education. Such research may not only shape the Assimilation Period, but may, as was the case with Kahneman and Tversky's research during the Post-Piagetian Period, define the Assimilation Period.

Based on the above, we contend that the next period of research, the Assimilation Period, may be defined as a renaissance period for psychological research in mathematics education. If not, then, at the very least, it is essential "for theories about mathematics education and cognitive psychology to recognize and incorporate achievements from the other domain of research" (Gillard et al. 2009, p. 13). As a caveat, we would add that such recognition and incorporation of achievements could be bidirectional.

3 Perspective III: Stochastics

Within Perspective III, we have chosen two themes to highlight: randomness and statistics. However, as discussed in the Preface to this perspective, it would be more accurate to declare the two themes of this perspective to be: perceived randomness and statistics.

Given the close connection between randomness and probability, one might expect a similar volume of research for the two topics. This, however, is not the case. The volume of research into probabilistic thinking, at least in the field of mathematics education, is greater than the volume of research investigating randomness and perceptions of randomness. However, this perspective houses some unique investigations into (perceptions of) randomness (Batanero et al.; Jolfaee et al.; Saldanha and Liu). These studies not only inform the existing research literature related to this topic, but, further, forge new threads of research by combining topics in probability (e.g., randomness) with other areas of research in mathematics education (e.g., gesture). Ultimately, these investigations may stimulate further research on (perceptions of) randomness and, thus, increase the volume of literature on this topic.

Statistics and probability are inextricably linked; yet, their relationship, in the field of mathematics education, is uncertain. In particular instances, major research syntheses combine probability and statistics research (e.g., Shaughnessy 1992) while, in other instances, the two areas are kept separate (e.g., Jones et al. 2007). What is certain, however, is that statistics is a major area of research in mathematics education. Indicators include, for example, conferences and journals solely dedicated to statistics education. However, the same is not true for probability—there is no [*Probability*] *Education Research Journal* or *Journal of* [*Probability*] *Education*. Additional indicators of statistics being a major area of research include the field's influence on other areas of research in mathematics education (e.g., simulation, inference, and modeling). In relation to Perspective III, such an influence upon research into probabilistic thinking is witnessed in this volume (e.g., English; Prodromou; Roth). Statistics education, then, provides an example of the potential for research investigating probabilistic thinking to become, one day, a major area of research in mathematics education (i.e., mainstream), which will influence related areas of research in mathematics education (e.g., statistics and others) and beyond.

4 Perspective IV: Mathematics Education

In considering the themes emerging from the mathematics education perspective, the last of the four main perspectives, we have chosen to comment on three. These themes are: the teaching and learning of topics in probability (and mathematics), areas of research in mathematics education, and methods used in mathematics education research.

In mathematics education, certain research investigating probabilistic thinking focuses on the teaching and learning of particular topics in (and associated with) probability. More specifically, some of the research (in this volume) investigates the teaching and learning of particular topics in probability; for example, the binomial formula (Sanchez and Landin), sample space (Paparistedomou), conditional probability (Huerta), the law of large numbers (Prediger and Schnell), counting methods (Maher and Ahluwalia), and chance (Nilsson). Research into the specific topics mentioned not only contributes to existing research for said particular topics, but, further, to existing research into probabilistic thinking, in general.

The second of the themes emerging from the mathematics education perspective was situating research into probabilistic thinking within other areas of research in mathematics education. Examples commented on and included in this volume are: the affective domain, that is, attitudes and beliefs (Nisbett and Williams), teachers' probabilistic knowledge (Mamolo and Zazkis; Eichler and Vogel; Batanero et al.), representation(s) (Maher and Ahluwalia; Sanchez and Landin), transfer (Watson) and cultural investigations (Sharma). Similar to the first theme discussed, research housed in this volume not only contributes to specific areas of research in mathematics education, but, further, to existing research into probabilistic thinking, in general.

Lastly (in terms of themes commented on for this perspective), methods used in mathematics education research were also represented in this volume. Methods, included and discussed to varying degrees, include: teaching experiments (Nilsson; Nisbett and Williams), teaching/learning arrangements (Prediger and Schnell), longitudinal studies (Maher and Ahluwalia; Prediger and Schnell), frameworks (Sanchez and Landin; Mooney, Langrall and Hertel), and simulation (Nisbett and Williams; Paparistedomou). Similar to the previous two themes, said research not only contributes, specifically, to the methods above, but, generally, to research investigating probabilistic thinking.

5 Final Perspectives

As we have detailed above, a number of different overarching themes have emerged from each of the four main perspectives in this volume. If, however, we had to declare one theme that ran through this entire volume, it would have to be the humanistic tradition in mathematics education (see Sriraman and Lee, this volume). From Borovcnik and Kapadia's historical account of probability and their detailing of puzzles and paradoxes to the historical treatment of randomness (Batanero et al.; Mooney, Langrall and Hertel; Saldanha and Liu) to individual notions and perceptions of randomness (Batanero et al.; Simin, Zazkis and Sinclair) to the question of whether man is an intuitive statistician (see, for example, Perspective II) to cultural influences (Sharma) elements of the humanistic tradition run deep in this volume.

Concluding our commentary on *Probabilistic Thinking: Presenting Plural Perspectives* in the forward looking spirit of the *Advances in Mathematics Education*

Series, we wish to make a few final comments regarding what we referred to as the movement of research investigating probabilistic thinking into mainstream mathematics education, which we have further recognized as a move from the Contemporary Research Period to the Assimilation Period.

Until now, research into probabilistic thinking in mathematics education has drawn heavily on research from other domains. As but one example, such research has relied substantially on the field of psychology and, in many ways, provided the foundation of this field. Consequently, research into probabilistic thinking in mathematics education has not, necessarily, been well situated within mathematics education literature—it has depended on the research from other fields. To illustrate our point, we presented a list of terms that we argued would, one day, be evaluated for their liberal usage, that is, used without the traditional academic scrutiny that would be found in other areas of research in mathematics education. The list of terms was long and included: beliefs, cognition, conceptions (and misconceptions), heuristics, intuition, knowledge, learning, modeling, reasoning, risk, stochastics, subjective probability, teaching, theory (or theories), thinking, and understanding. As evidenced in this volume, we are at the beginning of the end. For example, we consider the chapters from Nisbett and Williams and Sharma as exemplars of this point. Not only are these chapters well situated in the pertinent research literature, but, further, they are well situated in the research on attitudes and beliefs and culture (respectively) found and flourishing in the field of mathematics education.

We await the day when a two-way street is finally opened, that is, research into probabilistic thinking in the field of mathematics education is used to inform other major fields of research (e.g., psychology or philosophy). We consider the chapters of Abrahamson, Bennett, and Simin et al. as exemplars of research that could, potentially, have an impact on other fields of research.

Should the day come, when research investigating probabilistic thinking in the field of mathematics education explicitly influences other areas of research, we wish to note (at this point in time) that we will have moved beyond what we have called the Assimilation Period to whatever the "next" period will be called.

We hope that you enjoyed *Probabilistic Thinking: Presenting Plural Perspectives*. We further hope that each time you delve back into this volume you find another perspective, which will present probabilistic thinking in an ever greater context.

References

Chernoff, E. J. (2012). Recognizing revisitation of the representativeness heuristic: an analysis of answer key attributes. *ZDM. The International Journal on Mathematics Education*, *44*(7), 941–952. doi:10.1007/s11858-012-0435-9.

Gillard, E., Van Dooren, W., Schaeken, W., & Verschaffel, L. (2009). Dual-processes in the psychology of mathematics education and cognitive psychology. *Human Development*, *52*(2), 95–108.

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: the psychology of intuitive judgment*. New York: Cambridge University Press.

Jones, G. A., Langrall, C. W., & Mooney, E. S. (2007). Research in probability: responding to classroom realties. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 909–955). New York: Macmillan.

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

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

Kapadia, R., & Borovcnik, M. (1991). Looking forward. In R. Kapadia & M. Borovcnik (Eds.), *Chance encounters: probability in education* (pp. 255–256). Dordrecht: Kluwer.

Khovanova, T. (2011). Martin Gardner's mistake. arXiv:1102.0173v1 [math.PR] arXiv. http://arxiv.org/pdf/1102.0173.pdf.

Leron, U., & Hazzan, O. (2006). The rationality debate: application of cognitive psychology to mathematics education. *Educational Studies in Mathematics*, *62*(2), 105–126.

Leron, U., & Hazzan, O. (2009). Intuitive vs. analytical thinking: four perspectives. *Educational Studies in Mathematics*, *71*, 263–278.

Shaughnessy, J. M. (1992). Research in probability and statistics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 465–494). New York: Macmillan.

Tzur, R. (2011). Can dual processing theories of thinking inform conceptual learning in mathematics? *The Mathematics Enthusiast*, *8*(3), 597–636.

Vazsonyi, A. (1999). Which door has the Cadillac? *Decision Line*, *30*(1), 17–19. Retrieved from http://www.decisionsciences.org/DecisionLine/Vol30/30_1/vazs30_1.pdf.

Index

A

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0,

D

E

J

K

L

N

O