

Chapter 6: Research on Uncertainty

Dave Pratt

Institute of Education, University College London, UK

d.pratt@ioe.ac.uk +44 (0)1926 710185

Sibel Kazak

Pamukkale Üniversitesi, Turkey

skazak@pau.edu.tr +90 258 296 1025

Abstract We discuss research on the teaching and learning of uncertainty, with a particular emphasis on quantifiable aspects as might be represented by probability. We acknowledge earlier reviews of the field by integrating research, especially from the last ten years, with previous studies. In particular, we focus on three issues, which have become increasingly significant: 1. the realignment of previous work on heuristics and biases; 2. conceptual and experiential engagement with uncertainty; 3. adopting a modeling perspective on probability. The role of the teacher in shaping the learning environment in various critical ways emerges as a key finding. In the concluding section, we indicate promising directions for research, including the need for more exploratory research in new areas such as the role of modelling and carefully designed experiments to test hypotheses that are apparent from more established studies.

Keywords Probability; Heuristics; Biases; System 1; System 2; Modelling; Scaffolding; Dialogic thinking; Distribution; Sample Space

6.1 Introduction

The notion of uncertainty is a broad concept that includes phenomena that lie outside the domain of statistics, which focuses on uncertainty due to random variation, when it is often possible to make inferences and predictions. Within this subset of uncertainty, it is sometimes possible to measure how uncertain a phenomenon is, and we refer to this term as ‘probability’. Probability theory provides tools for expressing, quantifying and modelling uncertainty. This chapter focuses on research concerning those key ideas and issues in uncertainty and probability that are seen as conceptual links to statistics. We first give an overview of the previous reviews related to the topic and then introduce our approach to reviewing the research literature beyond those covered in the preceding documents.

There have been several edited books (e.g., Kapadia and Borovcnik 1991, Jones 2005, Chernoff and Sriraman 2014) and a number of major review chapters and reports on research in probability since probability and statistics started to become part of the mainstream school mathematics curricula in many countries. In his review in the Handbook of Research on Mathematics Teaching and Learning, Shaughnessy (1992) set the stage by addressing the absence of probability and statistics in school mathematics, particularly in the US prior to ‘The Curriculum and Evaluation Standards for School Mathematics’ (NCTM 1989). He then used philosophical and historical influences in the development of probability as the backdrop for research in probability and statistics. In Shaughnessy’s (1992) review, studies in the research literature were clustered in three main areas: 1. different types of thinking used in making an inference or judgment under uncertainty (i.e., heuristics, biases and misconceptions) that are identified and documented primarily within the psychology research tradition, such as the influential work of Daniel Kahneman and Amos Tversky in 1970s and 1980s; 2. development of concepts of probability in different age levels; 3. effects of

interventions, such as types of tasks, instructional approaches, and use of computer technology, on students' conceptions of probability.

Another important review of the existing research at that time by Borovcnik and Peard (1996) focussed on probabilistic thinking and teaching of probability in the school mathematics curriculum. Borovcnik and Peard shed light on what hindered learning of probability by making distinctions between probability and other mathematical concepts and similarly between probabilistic thinking and other types of thinking (logical and causal). They also described how the history of teaching probability evolved both in Europe and the US as probability and statistics became part of the school mathematics curricula in different countries. Then various didactical approaches aiming to enhance teaching probability were discussed.

In the Second Handbook of Research on Mathematics Teaching and Learning, Jones, Langrall and Mooney's (2007) chapter revealed how much progress had been made both in the treatment of probability in curriculum documents and the research tradition since Shaughnessy's (1992) review. One of the focuses of Jones et al. (2007) was the content and pedagogical insights of three curriculum documents from the US, the UK and Australia at different grade levels (elementary, middle and high school), which were published around the same time. In these curriculum documents, introduction of probability to students started at early grades in the elementary school while they began to focus on more advanced ideas in probability at the high school level. In terms of pedagogy, the tasks at the elementary grades were more in line with students' experiences in the way that they allowed students to test their intuitions and overcome their misconceptions. As students reached advanced levels, investigations and applications were emphasised. Simulations and modelling chance situations, which had been suggested in Shaughnessy's (1992) review, were seen to become part of the middle grades and high school curricula (Jones et al. 2007).

The research literature covered in Jones et al. (2007) also reflected these changes in mathematics curricula by focusing on various conceptual topics relevant to probability, such as chance and randomness, sample space, probability measurement (including conditional, theoretical and empirical probabilities) and cognitive models of probabilistic reasoning. Related to the research on teaching of probability, Jones et al. highlighted the contributions made about teachers' content knowledge, pedagogical content knowledge and knowledge of student cognition in probability. Another distinct topic raised in this review was the idea of probability literacy based on Gal's (2005) work and its implications for content and pedagogical approaches in probability instruction.

Bryant and Nunes (2012) provided a detailed report on research documenting children's difficulties in learning and reasoning about probability and recommendations for future research, particularly on methodological aspects. They argued that four ideas in probability were key to successful learning in probability: 1. understanding randomness and its consequences; 2. analysing the sample space; 3. quantifying probability as a ratio; 4. developing correlational reasoning which involved the coordination of the previous three ideas. The omission of aggregate thinking as relating to distribution – rather than just sample space - is surprising in the light of research reported below.

More recently Watson, Jones and Pratt (2013) took a critical approach when reviewing research studies into students' reasoning about uncertainty, many of which were mentioned in the previous reviews. Unlike others written mainly for researchers, the primary aim of this work was to elaborate the research-based findings to support pre-service and in-service teachers' understanding of the key issues about students' learning about probability. Given the technological tools that have become available in recent years, the use of simulations and modelling to help students develop reasoning about uncertainty was again emphasised in Watson et al.'s review, including for young students.

Our aim in this chapter is to focus on three issues that we see as key developments emerging out of the history of the topic as captured by previous reviews mentioned above. The first issue, the realignment of heuristics and biases, is chosen because the research on heuristics has been a major focus of research in the field for many decades and a recent publication makes it timely to reconsider that body of work. The second issue, conceptual and experiential engagement with uncertainty, gives an account of recent developments in the main effort of research in the field, some of which might in fact be influenced by the first issue. The third issue, adopting a modelling perspective on probability, emerges directly from considerable development in the use of technology for teaching and learning and also for researching students' ideas about uncertainty. As we introduce each of these three key issues in the following sections, we give an overview of the research in learning and teaching of probability over the last ten years and look forward to future research in this domain.

6.2 The Realignment of Heuristics and Biases

6.2.1 Introduction

We begin our review with discussion of an issue that has informed - some would say beset - research on probabilistic thinking for several decades. The issue in question was a particular focus of the review of research in probability and statistics, which is now more than two decades old (Shaughnessy 1992), and has led to an industry of research identifying misconceptions and correlating them in support of, or in contradiction to, the original work. We speak of course about the seminal work by Daniel Kahneman and Amos Tversky (for example, Kahneman, Slovic and Tversky 1982), which claimed to catalogue the biases inherent in heuristics that we all use to make judgements of chance. This research has recently achieved new currency because of Kahneman's publication on 'Thinking Fast and

Slow' (2011a), which has re-conceptualised the original research and, in doing so, has responded to some of the original criticisms. Kahneman's realignment of his own work on heuristics has implications for interpretation of the wealth of research on probabilistic thinking, especially as related to misconceptions, that has emanated from the original work in the 1970s and 1980s.

Our approach to discussing this key issue will be first to summarise the original work. This can be done briefly since there are many full accounts available elsewhere, not least in the review by Shaughnessy (1992). We will then discuss some of the criticisms that emerged over subsequent years. All of this will be preparation for a detailed account of Kahneman's fresh perspective on that work, followed by a discussion of whether the old criticisms still stand and implications for research in the field.

Kahneman and Tversky conducted a series of carefully designed psychological experiments where subjects were given tasks, either orally or in paper and pencil form. Kahneman and Tversky noted the errors that subjects made when their responses were compared with normative probabilistic or statistical solutions to the task. They identified a number of patterns in those errors and accounted for these patterns in terms of the subjects using rules of thumb, perhaps subconsciously, which they referred to as heuristics. Kahneman and Tversky explained how errors resulted from the bias, which was inherent in the heuristics being used.

As explained above, it would be inappropriate here to detail the huge array of heuristics identified, especially as each of the heuristics also has many variations and specific types. Nevertheless, some readers may wish to have a sense of the original work without needing to know all of that detail, so we will describe here two of the main heuristics identified by Kahneman and Tversky.

6.2.2 Two Heuristics from the Original Research and Recent Developments

When people use the representativeness heuristic, they judge the likelihood of an event according to how well the outcome experienced matches the system that generated the outcome or the population from which the outcome was drawn. The well-known gambler's fallacy might be accounted for by use of the representativeness heuristic. Thus, after observing six successive red numbers appear on the roulette wheel, the gambler might place his bet on the appearance of a black number (an approach referred to as negative recency). Kahneman and Tversky argue that the gambler might believe that the outcomes should match the sample space, which consists of an equal number of red and black numbers and so the judgement may have been made that a black number should appear in order to 'correct' the sequence of reds. The representativeness heuristic operates in the gambler's judgement by attempting to match the outcomes with the sample space.

Another situation in which the representativeness heuristic can lead to erroneous judgement is when a specific condition is regarded as more probable than a single general one, often referred to as the conjunction fallacy. For example, given a pen portrait description of Linda as a woman who is single, outspoken and very bright, and deeply concerned with issues of discrimination and social justice, it is not unusual for subjects to respond that Linda is more probably a bank teller and active in the feminist movement than that Linda is a bank teller.

Kahneman and Tversky argue that the representativeness heuristic will often provide correct judgments but, since representativeness does not allow for the vagaries of random chance, nor the laws of probability, there will be situations in which representativeness generates the wrong judgement, a systematic error that the authors refer to as 'bias'.

A second major heuristic identified by Kahneman and Tversky is 'availability'. People sometimes make a judgement about the chance of an event on the basis of how easily they are

able to evoke particular instances of the same or similar events. For example, the risk of a crash by the aeroplane in which you are travelling may seem disproportionately high (when compared to the frequency of recorded accidents) if there has been a recent widely reported tragic case of such an incident in which many people died. As with representativeness, the availability heuristic will often generate a correct judgement but how easily instances of an event can be evoked is highly sensitive to the salience of the event. The salience of an event is not generally related to its likelihood, which results in a bias inherent in the availability heuristic.

In the last ten years there has been further research on the trajectory over time of heuristics and biases and errors. Bennett (2014) studied 163 first year college students (though this group was divided into several treatments so the sample size for any one experiment was in the low thirties). The study found that students working on tasks inspired by the Monty Hall problem demonstrated a strong tendency for their decisions to be shaped by the ‘endowment effect’, an unwillingness to tempt fate by changing one’s mind about a decision in the light of further information, even when a rational decision according to probability theory would be to do so.

Chiesi and Primi (2009) investigated how the errors due to negative (and positive) recency developed or receded with age. They tested 23 primary school third graders, 25 primary school fifth graders and 35 college students. They found that, whereas positive recency (in which, for example, the gambler would bet on another red number at the roulette wheel after a sequence of red numbers) decreased with age, the negative recency effect was unaffected over time.

Kustos and Zelkowski (2013) examined misconceptions in probability tasks in the form of a survey consisting of open-ended structured questions for between 500 and 600 students across grades 7, 9, 11 and also those of 40 third year pre-service mathematics teachers. These

misconceptions included inter alia recency effects and representativeness, in other words some of the errors that arise, according to Kahneman and Tversky, from biases in heuristic thinking. They found that the recency effect and representativeness dissipated with age.

There is an evident discrepancy between the above two studies of how misconceptions arising from heuristic thinking develop. The large-scale study of Kustos and Zelkowski suggests that factors in the development of students in middle to high achieving schools in Alabama have a positive impact on the students' probabilistic reasoning. Although the researchers offer implications for teaching, these must be regarded as speculative as pedagogy was not investigated in the study. The smaller study of Chiesi and Primi took place in Italian public schools and it is entirely possible that factors impacting on the development of these students were very different. It is also possible that the sample size in this study was too small. Further research is needed before we can understand these conflicting results and it may be that a better theoretical understanding of heuristic thinking is needed before we can really predict how errors might be affected by schooling or age. A new theoretical understanding is perhaps now beginning to emerge and is discussed later in this section.

6.2.3 Criticisms of Kahneman and Tversky's Original Work on Heuristics

The main critic of the heuristics and biases approach of Kahneman and Tversky has been Gerd Gigerenzer. In his own work, Gigerenzer has advocated the use of natural frequencies instead of probabilities or proportions to communicate risk (e.g., Meder and Gigerenzer 2014). Bodemer, Meder and Gigerenzer (2014) demonstrated that people with a wide range of numeracy levels were less likely to interpret relative risk reductions in heart disease as absolute reductions when the baseline risks were presented in frequency format than when they were presented as percentages. We note however that Diaz and Batanero (2009) conducted a comparison of performance amongst 206 students, who took a test after a teaching unit on conditional probability, with a comparable group of 177 students, who took

the test before the course. They argue that detailed analysis of the types of errors apparent at different stages of a solution led them to teaching approaches that have demonstrated improvement in performance, even when probabilities rather than natural frequencies were used in conditional probability problems. (Specific cases where they did not find improvement are reported below.)

Gigerenzer (1991) has argued that some of the errors identified by Kahneman and Tversky disappear when the information is presented in a frequency format. Kahneman (1996) in turn has responded that their own studies supported the notion that presentation format impacted on the use of heuristics. However, he argued that this did not undermine the observation that subjects made systematic errors when presentations were not frequency based. Kahneman added that, though these errors might have been reduced, they did not disappear when the format was changed to one of natural frequencies, except perhaps in some very specific types of heuristic, such as the conjunction fallacy. Interestingly, in the Diaz and Batanero (2009) study, using probabilities rather than natural frequencies, the conjunction fallacy was one of the few errors that was resistant to improvement through their teaching methods.

Gigerenzer (1994) has also argued that there are difficulties with Kahneman and Tversky's focus on errors, which requires a normative position against which to judge the subjects' responses. They argue that there is fundamental disagreement amongst statisticians about the nature of probability, especially in relation to unique events, where a frequentist interpretation of probability does not apply. Of course, in many situations frequentist and subjective interpretations of probability converge and Kahneman (1996) pointed out that much of their historical work was not based around subjective probabilities. In fact, Gigerenzer's (1993) philosophical position regarded people's use of heuristics as rational acts, where decision-making apparatus has evolved so that decisions can be made when time and resources are limited. In his view, such apparatus rationally seeks out heuristic-based

methods of decision-making at the expense of accuracy and that these rational methods can be more accurate than formal methods. Hence, whereas Kahneman and Tversky have presented a fallible human who makes errors due to the use of inherently biased heuristics, Gigerenzer has offered a rational human who uses heuristics that are often accurate to make quick decisions on complex judgements of chance.

Perhaps, this theoretical difference goes to the heart of the prolonged dispute that has stretched across many publications, through the 1980s and 1990s. Although in his review, Shaughnessy (1992, p. 470) referred to the Kahneman and Tversky work as providing a theoretical framework for mathematics educators, one criticism of the work has been that it is in fact atheoretical. In response, Kahneman (1991) argued:

I take the distinctive feature of theory to be a commitment to completeness (within reason) and a consequent commitment to critical testing, in a specified domain of refutation, which is often quite narrow. (p. 143)

The difficulty for educationists lie in how to interpret Kahneman and Tversky's original work without a theoretical account of knowledge, thinking and learning. The catalogue of errors might be interpreted as suggesting that fallibility with respect to judgements of chance is integral to the human condition, which would be a bleak interpretation for those who hope to intervene in a student's understanding of probability. On the other hand, perhaps awareness of the bias in the use of heuristics, such as representativeness and availability, could be sensitised with the possible effect of improved judgements of chance. Clearly, as psychologists with an interest in decision-making, Kahneman and Tversky were not attempting to offer advice to educationists. Nevertheless the recent publication, 'Thinking Fast and Slow', by Kahneman (2011a), does situate the original research in a theoretical framework, which makes it possible to interpret the implications of the original work in new

ways, and perhaps sheds new light on the theoretical difference between Kahneman and Gigerenzer.

6.2.4 Heuristics as Part of System 1 and System 2 Thinking

Kahneman (2011a) has recently adopted dual process theory, and in particular the terminology of Stanovich and West (2000), to refer to System 1 thinking as automatic, quick and requiring little or no effort with no sense of voluntary control. In contrast, he stated that System 2 thinking is effortful, often involving complex computations, associated with agency, choice and concentration. To take one of Kahneman's examples, look at the following problem: 17×24 . System 1 tells you immediately that this is a multiplication problem (and may even allow you to estimate a rough answer). However, to compute the actual value requires the slow thinking of System 2. Loosely speaking, if System 1 were regarded as intuition, System 2 could be thought of as formal reasoning. Kahneman argued that much decision-making, and certainly that which involved the heuristics he had identified in his earlier work, operated at the automatic, largely subconscious, level of System 1, whereas the careful application of scientific theory and procedures to reach a decision demanded the effort of System 2. While System 1 by default was triggered automatically to make quick decisions, often with limited evidence, occasionally System 2 was activated when System 1 ran into trouble, such as when System 1 did not generate an answer, but System 2 required more time and resources. We can offer another illustrative example taken from a probability study by Kazak (2015). Consider a game in which two bags contain counters. One bag has 3 blue counters and 1 red counter. The other has 1 blue and 3 red counters. A player chooses one counter from each bag and wins if the colours match. Typically, the symmetry of the bags leads students to a swift judgement (System 1) to think the game is fair insofar as there appears to be an equal chance of winning. However, a careful calculation of the sample space (System 2) shows that the chances of winning and losing are not equal.

It is worth noting that Fischbein (first in his seminal work of 1975 and then through many subsequent publications) developed a substantial account of the part played in probabilistic thinking by our primary (unschooled) and secondary (systematically-trained) intuitions. The description of System 1 thinking seems to match rather well this account of primary and secondary intuitions.

It is interesting to note that Babai, Brecher, Stavy and Tirosh (2006), studying probabilistic reasoning, reported results that could be interpreted as supporting the operation of System 1 and System 2 in that reasoning. They studied the responses and response times of 68 16 and 17-year-old Israeli students to 20 'congruent' test items where the solution was expected to be in line with an intuitive response and 20 'incongruent' items in which the solution was regarded as counter-intuitive. They found not only that accurate responses were more prevalent amongst congruent items but also that correct responses to congruent items were quicker than correct responses to incongruent items. This finding is consistent with System 1 finding immediate solutions to the congruent items but System 2 needing to find a more effortful solution to the incongruent items.

System 1 cannot be switched off (Stanovich and West 2000), so training System 2 to be less accepting of System 1 when System 1 readily finds a solution may become the focus for educationists. In Fischbein's terms, this could be one focus for promoting secondary intuitions. This is especially important for teachers and researchers of probabilistic thinking, identified by Kahneman as a conceptual field where System 1 often uses non-stochastic intuitions to respond to uncertainty.

At one level, we might recognise here Gigerenzer's portrayal of an evolved system that allows the heuristics (of System 1) to operate much of the time, perhaps forgoing accuracy (of System 2), for the benefit of speed (of System 1). Kahneman identifies a rich host of mechanisms that System 1 uses in order to reach quick answers to questions. For example, he

claims that one technique is to substitute an easier question than the one actually posed. According to Kahneman, substitution is in fact a particularly prevalent cause of heuristic errors in the field of probability and statistics. For example, System 1 cannot correlate information about baseline frequencies alongside intuitions about resemblance and so the representativeness heuristic tends to determine the decision. According to Kahneman, faced with a question about likelihood, System 1 substitutes a simpler question about resemblance. Another example is apparent when System 1 substitutes a question about the frequency of an event with a question about how easily similar instances come to mind, with the consequence that the availability heuristic tends to determine the answer.

Chernoff (2012) demonstrated the use of attribution substitution in probabilistic reasoning amongst 59 pre-service elementary and middle-school teachers. In an unusual variation on the tasks typically used to test for representativeness, the subjects were asked which of two answer keys (A C C B D C A A D B or C C C B B B B B B B) was least likely to be the answer key for a 10 question multiple choice maths quiz, each question having four possible responses. They were also asked for an explanation. (An answer key is the coded list of correct responses.) Chernoff concluded that certain individuals, when presented with one question, possibly unknowingly answered a different question, substituting a variety of heuristic attributes, such as ‘most resembling’ in place of ‘most likely’.

In contrast to Gigerenzer’s emphasis on the rationality of people’s use of heuristics, Kahneman’s focus remains on how people’s reliance on System 1 leads to systematic errors.

6.2.5 Implications of System 1 and System 2 for Probabilistic Thinking

One of System 1’s techniques for making quick decisions is to readily draw causal inferences from the evidence immediately available. When presented with data, System 1 will begin to observe patterns and form impressions as possible causal explanations. System 2 typically

accepts these explanations. This accounts for how we mistakenly see patterns in random behaviour, design in arbitrary events, intention in the accidental. According to Kahneman, this technique of System 1 explains why people, when presented with randomly generated data, use heuristics to predict how the sequence will extend. This attribute of System 1 also provides an account of why people confuse association with causation, attributing causality to patterns in data that might have no causal connection in reality:

people are prone to apply causal thinking inappropriately to situations that require statistical reasoning. Statistical thinking derives conclusions about individual cases from properties of categories and ensembles. Unfortunately, System 1 does not have the capability for this mode of reasoning; System 2 can learn to think statistically, but few people receive the necessary training. (Kahneman 2011a, p. 77)

In the Diaz and Batanero (2009) study, participants often confused causality and conditionality, and they typically assumed that the likelihood of an event could not be affected by the likelihood of an event that has already happened. These errors were resistant to improvement through their teaching methods. Perhaps because System 1 searches for causations, there is a tendency to account for conditional relationships as if they were causal with time dependence.

The difficulty people have in recognising a situation as amenable to a statistical interpretation has been well documented. Konold (1989) referred to people's tendency to focus on what happened, rather than on strategic probabilistic approaches, as the 'outcome' approach. Thus, focussing on outcomes, System 1 might easily infer causations even when the patterns noticed are explained merely by the vagaries of chance.

Lecoutre, Rovira, Lecoutre and Poitevineau (2006) investigated how 20 grade 3 pupils, 20 psychology researchers and 20 mathematics researchers, all based in Rouen in France,

decided on whether given situations might involve randomness. 16 items were presented on cards and varied according to whether: 1. the items were events from everyday life experiences or a repeatable process that might involve random variation; 2. the items addressed the subject as 'you' or not; 3. the possible outcomes were equally likely or asymmetric. First the subjects were asked to categorise the 16 items for themselves and then they were asked which items involved randomness. The researchers concluded that subjects decided randomness was involved when they could recognise probabilistic reasoning, for example by being able to compute a probability, making probability rather than randomness the foundational idea. Subjects decided randomness was not involved when they thought determinism played the larger part or when causal factors could be identified. Since System 1, according to Kahneman, is constantly searching for causal patterns, it is perhaps not surprising that the possibilities for a stochastic approach tend to be ignored and people demonstrate the outcome approach.

Smith and Hjalmarson (2013) examined 32 pre-service mathematics teachers' conceptions of random processes with respect to the traditional game of 'rock, paper, scissors'. Teachers found it difficult to reconcile equality of winning outcomes for each player with the human interference apparent when choosing how to place their fingers in the game. System 1 all too easily recognises the human element as a causation but this conflicts with notions of fairness, often associated with randomness (Pratt 2000, Papanastasiou, Noss and Pratt 2008, Papanastasiou 2014). The pre-service teachers did ultimately decide that the outcomes were not generated randomly. The researchers concluded that understanding of the nature of randomness developed during their instructional sequence as a result of making the generating process explicit and focussing on whether that constituted random generation or not.

In their early work, Kahneman and Tversky introduced the so-called ‘Law of Small Numbers’ to describe how people behave as if the Law of Large Numbers applies to short sequences as well. There is a tendency to underestimate the need for samples with large numbers in order to draw reliable inferences. Kahneman now explains this in terms of System 1. When samples are small, apparent patterns can be identified simply because extreme results are more likely to happen than if the sample were large, and System 1 tends to attribute causal explanations to those patterns.

More broadly, Kahneman argues that we easily think associatively, metaphorically and causally and these styles of thinking are more suited to System 1 than is statistical thinking. It remains an open question as to whether educationists will be able to find ways of training or educating their students such that System 2 would be less accepting of System 1 answers in identifiable scenarios. There is some evidence, presented below, to suggest that this may be possible. At the point that System 2 is required to affirm System 1’s answer, it might be possible to teach System 2 to be less easily convinced in certain scenarios that capture typical probabilistic and statistical situations.

6.2.6 Intervention Studies on Heuristics and Biases

Below, we focus on intervention studies, which may suggest pedagogic methods to address the difficulties that seem to be generated through System 1 thinking.

Fast (2007) conducted a study of 54 female Zimbabwean students. A test, consisting of questions of the sort used by previous researchers to identify misconceptions, was administered. The students were found to make errors in their responses that were consistent with representativeness, availability and other heuristics. Source analogues were constructed and offered to the students through interviews. These analogues were designed to be structurally similar to the initial test items but were intended to generate normative responses

and so be the basis of knowledge reconstruction, which was evaluated as generally successful. For example, a source analogue might pose a similar problem to that posed by the original test item but with the situation amended so that the numbers were more extreme. Thus, in the original test, subjects were asked whether a sports team, thought to be better performer, would be more likely to win against a supposedly inferior team in a playoff based on 5 matches or 9 matches. In contrast, the analogue question compared a single play-off match with a 5-game playoff. The intention was that subjects would be able to use common sense to find the correct response to the analogue question, and then recognise its structural similarity with the original test item. A delayed post-test suggested that the analogues continued to provide anchors for normative thinking one month later. The process of knowledge reconstruction was seen as critical. Even though this research was based on a fairly small and specific group, the above intervention raises the question whether, in Kahneman's terminology, the use of analogues might offer a bridge towards normative thinking by sensitising System 2 to a set of scenarios in which System 1's automatic and quick response might otherwise be problematic.

Another approach has been demonstrated over several years, in the work of Pratt (2000) and Pratt and Noss (2002, 2010), where the intervention was based around children mending computer-based 'gadgets', virtual simulations of everyday random generators, whose configuration could be edited to make them work properly. These 10 to 11-year-old children tended not to recognise that attributes of randomness in the short-term (e.g., unpredictability, lack of control over the outcomes, irregularity in results) differed from attributes of randomness in the long term, at least from the aggregated perspective (where relative frequencies become predictable and aggregated results have a regularity to them). From the Kahneman perspective, these children's System 1 heuristic thinking appeared to suggest that, when chance was operating, it was just a matter of luck. By working with the gadgets, the

children gradually became aware of patterns in the aggregated view over the long term. Pratt and Noss (2010) concluded that the key elements in the intervention design were: 1. enabling the testing by children of their personal conjectures; 2. seeking to enhance the explanatory power of knowledge that might offer a route to normalised knowledge; 3. constructing a task design that would be seen by the children as purposeful and allow them to appreciate the power of the mathematical idea of distribution; 4. designing a representation of distribution that could be initially used as a control point by the children and subsequently become a representation with predictive power. These design constructs perhaps offer some further insight into what might be needed in order to sensitise System 2 to the need to distinguish between scenarios with small versus those with large numbers.

Paparistodemou et al. (2008) also used a computer-based microworld to study 23 5 to 8-year-old children's ideas about fairness. The children were challenged to build a lottery machine by arranging a spatial configuration of red and blue balls, off which a small white ball would bounce. When the white ball hit a red ball, a character called the 'space kid' moved in one direction and when a blue ball was hit, the space kid moved in the opposite direction. The aim was to keep the space kid near to his starting position. Some of the children's configurations exploited symmetry so that in effect the white ball bounced in turns from red to blue and back to red. Others exploited random bouncing so that it was impossible to predict which colour would be hit next. These two approaches were associated with deterministic and stochastic strategies respectively. By placing emphasis on fairness in an expressive environment, the children were able to imagine fairness not only in terms of turn taking but also in terms of the vagaries of chance. The design constructs listed above (Pratt and Noss 2010) seem to apply to this study as well, especially with respect to 1, 2 and 3. Kahneman might argue that the approach used in the Paparistodemou et al. (2008) provides

System 2 with new possibilities for how fairness, when detected by System 1, might be interpreted.

An intervention by Canada (2006) might be seen as analogous to that by Papanastasiou et al. but with respect to variation in probability situations. Canada's use of hands-on activities, supplemented by small-group and whole-class discussion of variation, with pre-service teachers may have enhanced their appreciation of how variation plays a role in statistical thinking.

Another approach that might enhance students' System 2 recognition of the possible weakness in System 1's proposed solution is to improve teachers' pedagogical knowledge of the types of reasoning students might use. Such a development might alert teachers to the need to artificially engage their students' System 2 thinking, with the aspiration that, after sufficient training, their students might begin to recognise such situations for themselves. There appears at least to be a deficit in teachers' knowledge about students' probabilistic reasoning. In an interesting study, Watson and Callingham (2013) examined the probabilistic reasoning of 247 students, mostly from years 7 to 11, and compared that to how their 26 teachers recognised their students' reasoning. Some of the students' reasoning was unfamiliar to the teachers suggesting that there might be value in finding ways of enhancing the teachers' pedagogical knowledge in this area.

6.2.7 Discussion

In this section, we have considered a key issue that has emerged in research on heuristics for making judgements of chance because of Kahneman's (2011a) recent publication on two reasoning systems. Our perspective is that this issue is very important for researchers in statistics education, who are interested in randomness and probabilistic thinking, because dual process theory allows us to interpret research in the field in new ways.

The debate between Kahneman and Gigerenzer continues. In ‘Thinking Fast and Slow’, there are several references by Kahneman to Gigerenzer’s criticisms. In fact, Kahneman takes the opportunity to criticise Gigerenzer’s notion of fast and frugal heuristics on the basis that, in Kahneman’s view, there is no imperative for the brain with its massive processing power to be frugal. Meanwhile, Gigerenzer (2012) has described how methods of making rational choices are inefficient when key factors influencing the decision are unknown. For more recent developments in this ongoing debate, see Kahneman (2011b) and Gigerenzer (2014), where there is a chapter on revolutionising schools through a risk-based curriculum. This emphasis on a risk-based curriculum is in line with Fischbein and other researchers who have argued for many years that the curriculum is predominately anchored in deterministic reasoning (deduction, proof, algorithms) and has historically ignored stochastic reasoning under uncertainty (statistical thinking).

Overall, we have summarised Kahneman’s application of dual process theory to his research and we have re-interpreted recent research in those terms as a means to offer insight into its implications. Nevertheless, we acknowledge that it is perhaps too early to offer a critical evaluation of the realignment of the heuristics research as proposed by Kahneman beyond the discussion above about implications. In subsequent sections, we address other issues which we see as recent key developments in research on probabilistic thinking and, although the emphasis will move away from Kahneman’s ‘Thinking Fast and Slow’, we invite the reader to attempt to interpret this research from that perspective, which might indeed yield further insights.

6.3 Conceptual and Experiential Engagement with Uncertainty

6.3.1 Introduction

Probability is a means to quantify uncertainty in random processes. Understanding how the concept of probability historically developed provides a perspective for interpreting current research results on students' conceptions of probability. One important aspect of probability that appeared in the mid 1600s is its duality (Hacking 1975, Weisburg 2014). The dual notion of probability implies that on the one hand probability is considered as degree of belief (subjective notion) and on the other hand it refers to stable frequencies in the long run (objective notion). Another approach to estimating probability, especially in games of chance, involves a priori method that requires an assumption of equiprobability.

Accordingly, there are three main schools of thought in probability theory that have different conceptions/interpretations of probability. From the classical view, the probability of an event is a ratio of the number of favoured outcomes to the total number of equally likely outcomes. In the frequentist view, the probability of an event is defined as the limit of the relative frequency of the observed outcomes as the number of trials increases indefinitely when a random experiment is repeated under identical conditions. The subjective interpretation of probability emphasizes personal probability relative to our background knowledge and beliefs.

The ongoing historical debates about different interpretations of probability have been also reflected in school curricula and in teaching of probability, such as theoretical, empirical, and subjective probabilities (see Jones et al. 2007). While existing research on heuristics revealed the inconsistencies between students' informal conceptions of probability and formal theory of probability (see earlier section on heuristic thinking), many recent research studies investigated how students' probabilistic conceptions developed and the ways to support them. In this section we focus on this body of research. The first part focuses on research that is primarily about students' understanding, though we suggest implications for teaching.

Subsequent parts consider how such understandings might be influenced by teachers, through the tasks they choose, their pedagogic approaches, and the tools they offer to their students.

6.3.2 Recent Research on Conceptual Development

Given the historical development of various meanings of probability, the concept of probability has a slippery aspect. Furthermore, the seminal work by Piaget and Inhelder (1951) and Fischbein (1975) offered a starting point for much research, reviewed in detail elsewhere (Shaughnessy 1992, Borovcnik and Peard 1996), that showed how the learning of probability is troublesome. More recently, several researchers have been particularly interested in the development of these conceptions from a variety of theoretical perspectives. Below we first summarise that work and then, in the final subsection, we draw together the implications for teaching.

Kafoussi's (2004) study focussed on the early development of quantitative reasoning about the likelihood of chance events during a classroom teaching experiment in a kindergarten. Individual interviews with children were conducted before and after the teaching experiment. Responses of the 5-year-old children during the pre-interviews tended to rely on subjective beliefs when judging the likelihood of given events. While children were able to identify all possible outcomes of a single-stage chance experiment, they could not give a complete answer for a two-stage experiment. They also seemed to have difficulties in comparing the likelihood of events when the task involved comparing of numbers of objects in a box rather than sizes of sections on a spinner. The post-interview results suggested considerable progress in children's probabilistic thinking showing a shift from subjective conceptions to a 'naive quantitative reasoning' as in Jones, Langrall, Thornton and Mogill's framework (1997, p. 121). Kafoussi argued that 5-year-olds' conceptual development was fostered during the teaching experiment as they began to: 1. discuss what counted as 'different' outcomes in a two-stage experiment; 2. consider the empirical results from an experiment as a solution to a

probability problem; 3. predict the results of a probability situation with equiprobable outcomes without conducting an actual experiment.

Prediger (2008) reported on a clinical interview study with ten pairs of 10–11-year-old children by focussing on their individual conceptions of chance situations in a game context before any probability instruction at school. Prediger found three categories of conceptions when children were explaining or justifying the outcomes or their predictions: everyday conceptions, empirical conceptions and theoretical conceptions. She was cautious about simply making a correspondence between these individual conceptions and three interpretations of probability (subjective, frequentist and classical). She suggested that some of these student conceptions could later be developed into a subjective conception of probability or a frequentist conception. However, one pair of students seemed to develop a notion of a classical interpretation of probability when talking about the number of different ways to find the sum of two die. Apart from this one example where the students had a learning trajectory progressing from everyday conceptions to the classical conception of probability, the other pairs seemed to move back and forth between different conceptions.

Prediger however did not treat the individual conceptions that were not theoretically sound as misconceptions in a traditional sense (i.e. (mis)conceptions to be substituted by the mathematically appropriate ones). Using the approach of horizontal development in the conceptual change research tradition, she considered students' everyday conceptions "as concurrent conceptions which co-exist with newly developed mathematical conceptions even in the long run" (Prediger 2008, p. 142). Similar to previous findings (Konold, Pollatsek, Well, Lohmeier, & Lipson, 1993; Pratt & Noss, 2002), the students' fluctuations between different conceptions during the task suggested that an individual might hold a range of views (from informal to formal) at the same time and use different ones depending on how they perceived the stochastic situation or what they paid attention to (single outcome vs. long run

or short term vs. long term contexts). The horizontal view suggested a complementary perspective to the vertical view of conceptual change focussing on transformation of misconceptions to mathematical conceptions. Adopting this approach to conceptual development in probability seemed to provide a valuable perspective on ‘typical’ persevering misconceptions and how to re-conceptualise them to help learners.

Furthermore, Schnell and Prediger (2012) applied the vertical and horizontal conceptual change approach to the development of students’ conceptions of the empirical law of large numbers. However, their main focus in this paper was on the theoretical contribution of their fine-grained method for analysing the microprocesses of constructing conceptions by using a notion of ‘construct’ as the unit of analysis and of building links among them as a webbing of constructs. By microprocesses, they referred to moving from an initial construct to an advanced one or changing the function of a construct as new relations between constructs were formed. Schnell and Prediger argued these microprocesses would contribute to the vertical and horizontal conceptual changes, suggesting the possibility of a successful trajectory from a ‘haphazard’ view of changes in the chance outcomes to a stabilized view of patterns in the long-term context.

As shown in previous research on heuristics, students often come to classrooms with alternative conceptions of probability. Teachers need to be aware of these different interpretations of probability for helping learners develop the formal ideas. From this perspective, the study of Liu and Thompson (2007), focusing on teachers’ understandings of probability on various tasks, is of importance. Research was conducted with eight high school teachers participating in an 8-week seminar on teaching and learning of probability and statistics with deeper understanding from a constructivist perspective. Liu and Thompson focused on teachers’ ‘stochastic conception of probability’ which they aligned to the frequentist view; in contrast, they argued that a ‘relative proportion conception of probability’

can sometimes be drawn upon without consideration of a repeatable stochastic process. Some other non-stochastic interpretations of probability, observed in teachers' responses and discussions, seemed to resemble those that students often have, for example: 1. the outcome approach (Konold 1989); 2. reduction of sample space for a probabilistic event (i.e. given that either an event will happen or it will not happen, the probability is either 1 or 0); 3. the principle of indifference approach to probability (i.e. the probability is 50% because an event may happen or not). Liu and Thompson argued that these non-stochastic interpretations would actually depend on how people conceived the given situation.

6.3.3 The Impact of Task Design on Conceptual Understanding of Probability

Conceptual development of probabilistic ideas is, of course, shaped by experience. For example, according to Ainley, Pratt and Hansen (2006), students' conceptual understanding of the utility of a probabilistic idea is connected with their sense of the purposefulness of the task in which they are engaged. In pedagogic situations, tasks set by the teacher can sometimes seem artificial, lacking purpose or relevance from the perspective of the student, perhaps because the teacher is very aware of their responsibility to teach the syllabus. The challenge, and it is recognised as non-trivial, is to create tasks that are seen as purposeful by the student but result in the student gaining appreciation of how the statistical idea is powerful in helping them to complete the task.

An example lies in Pratt's (2000) study of children configuring computer simulations of random generators such as coins, spinners and dice, referred to as gadgets. The children found the task of trying to make the gadgets work properly purposeful and it led inexorably to them gaining a sense of how a probability distribution, contextualised in this study as the workings box of the gadget, had the power to predict aggregated outcomes in the long term but not in the short term. More generally, Ainley et al. suggested a range of heuristics for designing tasks that are likely to connect purpose and utility; tasks might: 1. have an explicit

end product; 2. involve making something for another audience to use; 3. contain opportunities for pupils to make meaningful decisions.

6.3.4 Scaffolding and Dialogic Thinking

As seen in the previous sections, misconceptions or biases that hinder students' probabilistic thinking are well documented. There are a few research studies examining how the pedagogic approach of the teacher might facilitate learning of probability.

Corter and Zahner (2007) initially worked with 26 graduate students in an introductory statistics course to examine the use of external visual representations in probability problem solving. Each participant was asked to solve eight probability problems using a structured interview protocol. This exploratory study indicated that students used a variety of visual representations and that the appropriate ones tended to facilitate students' problem solving. Zahner and Corter (2010) further researched the role of the external visual representations on solving probability problems (such as what kinds of representations were used for different problems, how and when) with another 34 graduate students. The interview-based research suggested that certain representations used spontaneously by the students helped them perform better in solving particular problems compared to those not using any. Selecting and using appropriate external representations in presented problems seemed to be an important part of the problem solving process in this study.

Ruthven and Hofmann (2013) described the development of a probability module for early secondary school using classroom based design research. A distinctive feature of this module was its pedagogical approach that was based on prior research on effective ways of teaching mathematics and science, especially in the UK context. This pedagogical intervention involved a teaching approach where students were encouraged to express their ideas, give explicit reasons for their thinking and take different perspectives, an approach termed

‘dialogic’ (see Mercer and Sams 2006). Dialogic talk used in small group work and whole class discussions during the activities became a tool that helped students move from their informal ideas about probability, including some of those heuristics and biases mentioned above (mainly used in System 1 thinking mode) to formal probabilistic reasoning (i.e. System 2). Further evidence from Kazak, Wegerif and Fujita (2015a), working with groups of 10 to 12-year-old children, supported the idea that scaffolding for dialogue as well as for content, alongside the use of technological tools, helped to generate breakthroughs in probabilistic thinking.

Kazak, Wegerif and Fujita (2015b) explored whether an analysis of two 12 year-old students’ activity based on dialogic theory might offer new insights compared to a Piagetian or Vygotskian analysis. The students were exploring the fairness of a variety of chance games, which they played manually but also built in *TinkerPlots 2.0* software (Konold and Miller 2011, <http://www.tinkerplots.com/>). The researchers found that the Piagetian and Vygotskian analyses ignored what for most viewers of the activity was a very obvious phenomenon. The recordings of the activity showed how the students engaged in laughter, sometimes quite raucous, a phenomenon ignored by Piaget and Vygotsky, but of great interest to Bakhtin, whose work inspired the dialogic approach (Bakhtin, 1986). According to the authors, laughter creates space and openness for participants to switch perspective, and so to take the point of view of the other. More generally, they argued that switching perspective was facilitated by the good relationship between the participants, including the teacher, good humour being one indicator of such a relationship.

6.3.5 The Role of Technology

In considering how teachers might influence students’ understanding, we have so far considered recent research on task design, scaffolding through external visualisations and dialogic approaches. We now consider the tools, in particular technological tools, that they

might offer their students. Research continues to suggest that certain types of technology, used within carefully designed situations, can offer opportunities for probabilistic learning that stretch beyond those available in everyday experience. Biehler, Ben-Zvi, Bakker and Makar (2013) provided a recent review on such possibilities at school level. That review emphasised in conclusion some recurrent important points in the design of the learning environment that incorporates the use of technology: 1. skill was needed, by the user or the teacher, to know when it was appropriate to adopt a hands-on approach and when software might help; 2. one key feature of modern pedagogic statistical software laid in its dynamic, visual and personal nature; 3. one key focus needed to be on reasoning with aggregates; 4. the tension between adopting the power that technology offered and the time it took to learn and adapt to that technology needed to be addressed. With our specific focus on probability, we elaborate below a few research-based studies which we believe add to the above list of specific proposals for the design of a probabilistic learning environment but which were not detailed in that broader review.

Earlier, we mentioned Pratt's (2000) study in which 10 and 11-year-old children began to acknowledge that there were regularities in the aggregated results of random processes even though the same could not be said in the short term. In the previous section on heuristics, we set out the design constructs that, according to Pratt and Noss (2010), supported the development of those insights. Apart from those aspects of the design, it is clear that the technological environment provided the opportunity to gather artificial experience of the long term because the technology offered systematic feedback, quickly and repeatedly, which would not usually have been the case in everyday experience.

Similar results have been reported by Lee and Lee (2009), when children cheered for a chosen colour to be the most frequent in repetitions of computer simulated draws of marbles from a bag, only to find that the result was rather predictable, except in a short run. They

concluded that, in similar conditions to those reported by Pratt and Noss, students began to notice variability in small samples and regularity in large samples. When it came to interpreting the impact of adding some new data in small samples (more change/instability) vs. in large samples (less change, more stability) in the computer simulation results, other semiotic tools, such as the use of metaphors in combination with technology (Abrahamson, Gutiérrez, & Baddorf, 2012), helped students make sense of the visual phenomenon.

Ben Zvi, Aridor, Makar and Bakker (2012) studied how children aged 10-11 years expressed uncertainty while they conducted informal investigations of data. The students used *TinkerPlots 2.0* to make informal inferences on samples of data where the sample size was gradually increased. The students initially oscillated between deterministic and relativistic statements. Eventually, a basic probabilistic language began to emerge. The authors concluded that more sophisticated inference-making was encouraged by attending to students' expressions of uncertainty when making judgements about trends in data.

Abrahamson, Berland, Shapiro, Unterman and Wilensky (2006) proposed an additional role for the computer. The authors of the paper discovered conflicts in their interpretations of a computer simulation in which three boxes were randomly coloured green or blue. A single run resulted in any one of eight possible configurations, called keys (for example, green, green, blue is one key). The authors happily ran the simulation without disagreement. When the authors began to create probabilistic models of the situation, they discovered their apparent agreement was not founded on the same epistemological assumptions. It was possible to model either the length of a run of repeated guesses until a specific key appeared or the frequency of a particular key in various size samples of guesses. The authors found it difficult to agree on how the first model failed to generate the expected bell-shaped curve, a disagreement that was only resolved when the authors had had the opportunity to program the situations, were confident that the program was bug free and had corrected any errors in

thinking through discussion. Programming on the computer was for them a necessary step to expose and critique underlying assumptions and models, differences, which had not been apparent from simply running a prepared simulation. Chaput, Girard and Henry (2008) made a similar point about modelling, which has some commonalities with programming insofar as both require the learner to express their ideas about what is being programmed or modelled. They argued that the use of modelling in statistics education is a delicate process because of the problematic epistemological basis of probability. They contended that the advantage of using computers resides not so much in their power and efficiency as in the analysis of random situations that needs to be done in order to design the model and translate that design into computer instructions.

In a sense, programming and discussion in Abrahamson's reflective article above acted to bridge across the differing probabilistic assumptions that the authors had held. Abrahamson and Wilensky (2007) reported how the design of pedagogical situations, including the use of technology, supported students to bridge intuitively, cognitively or historically conflicting ideas in probability. They referred to these conflicting ideas as being at opposite poles of a learning axis. They set out to design bridging tools that were intentionally ambiguous with respect to these extremes. These tools were presented as part of a broader learning environment, designed to stimulate engagement with and argumentation about the epistemological ambiguity. There is a connection here in how Abrahamson and Wilensky exploited ambiguity to set up cognitive conflict, subsequently resolved through discussion, and how Pratt and Noss (2010) referred to blurring control and representation in the way that the computer-based simulations were configured and used.

In summary, we might ask what have we learned about the role of technology in the teaching and learning of probability to add to the findings in the Biehler et al. review (2013). Certainly there is support (Lee and Lee 2009, Pratt 2000) for the idea that extended experience with the

virtual, repeatable and artificial experience offered by some technological environments can contribute to a focus on aggregate thinking called for in that review, with the result that students can begin to distinguish between variability in the short term and regularity in the long term. In addition, there is growing evidence (Abrahamson et al. 2006) that programming models might for some clarify epistemological distinctions in probability. Biehler et al. highlighted the concern that in some situations teachers might judge that adopting technological approaches is more time consuming than is warranted by the benefits that accrue and this could be a view taken by some teachers with respect to programming. The development of bridging tools (Abrahamson and Wilensky 2007) that have a degree of ambiguity with respect to contrasting epistemologies might offer a similar role to programming and be less time consuming for the student.

6.3.6 Discussion

In the first section of this chapter, we summarised the research on heuristics and biases and reviewed recent developments in theory that linked that earlier work to System 1 and System 2 thinking. According to Kahneman's account, System 1 thinking is relatively automatic and is best controlled by careful training of System 2. In the current section, we set out to review recent research to build on earlier reviews about how that might best be done.

What is clear from this review is the critical role played by teachers. Examples of this, cross referenced to the literature drawn on in this section, are:

1. offering more empirical hands-on experience of random variation (Biehler et al. 2013);
2. the artful selection of digital tools and other types of external representations (Pratt 2000, Zahner and Corter 2007, Lee and Lee 2009, Biehler et al. 2013);
3. focussing such experience on prediction to tease out what counts as different outcomes (Kafoussi 2004);

4. recognising the complexity of different epistemologies of probability and helping students to bridge the apparent discrepancies through programming or specially designed tools (Abrahamson et al. 2006, Abrahamson and Wilensky 2007, Liu and Thompson 2007, Prediger 2008);
5. acknowledging the importance of task design, since the situation in which random variation is met influences how people think about probability and because purposeful tasks can, if carefully designed, lead to a sense of the power of the probabilistic concepts (Ainley et al. 2006);
6. offering opportunities for students to express their ideas with their peers and through technology so that ideas can be negotiated and perhaps converge (Ruthven and Hoffman 2006, Ben Zvi et al. 2012, Kazak et al. 2015a, Kazak et al. 2015b).

Some of the above ways in which teachers might support learning of probability are especially suited to an approach in which probability is seen as a key part of creating or exploring models of situations that are amenable to a statistical interpretation. Modelling is therefore the focus of the next section in this chapter.

6.4 Adopting a Modelling Perspective on Probability

6.4.1 Introduction

One of the striking developments in recent research on probability (and its connections to statistics more generally) is the increased emphasis on modelling. Models have always been a key element of statistics as a discipline in the way that they describe data probabilistically (for example in the form of probability distributions or analytical methods such as analysis of variance). According to Wild and Pfannkuch (1999), modelling is also an important component of statistical reasoning. The emergence of modelling in teaching and learning has

no doubt been driven by the increasing access to technology and improved software, especially that aimed at learners. Modelling appears to have the potential to facilitate the methods by which teachers can support learners, as listed in the previous subsection. Indeed, modelling promises to offer a connection between data and probability (Konold and Kazak 2008) that is meaningful to learners and may provide an approach that enables learners to appreciate the power of probability, at a time when dice and card games have become less of a focus of play for the younger generation than in the past.

Modelling approaches tend to place emphasis simultaneously on data and uncertainty.

Models can be developed to fit real data but the fit will not be exact, requiring a probabilistic element to the model in order to account for the variation in the data. Computational models can be executed to generate virtual data, which may approximately reflect the real data if the model was a good one.

Theoretical distributions and sample spaces can be thought of as models and so we begin this section by considering research in these areas. Subsequently, we consider research that addresses explicitly how a modelling perspective on probability might influence understanding (see Chapter 7).

6.4.2 Understanding Empirical and Theoretical Distributions

In their earlier review of student learning of probability, Jones et al. (2007) commented that, in view of its importance in curricula, it was surprising that at that time there was little research on student conceptions of experimental probability. They did quote limited evidence about the difficulty students experience in making links between the sample space of a random generator and outcomes actually generated. They also noted the proclivity for students not to realise the connection with the use of large samples until they were able to spend extended periods working with simulations that allowed the use of samples of any size.

There has since then been further research on students' understanding of theoretical and empirical distributions.

Ireland and Watson (2009), researching 10–12 year-old students, concluded that it was insufficient for educators to focus on the calculation of theoretical probabilities and the observation of experimental outcomes. According to their study, the connection between experimental and theoretical probability needed to be taught and experienced explicitly, by encouraging the creation of new correct probabilistic intuitions, the prediction of outcomes, the performance of experiments and the evaluation of outcomes as advocated by Fischbein (1975).

More recently, English and Watson (2016) conducted such a teaching experiment on 91 9 and 10-year-olds, who tossed one and two coins, explored relative frequencies through graphing in *TinkerPlots 2.0*, which they also used to simulate large scale tossing of coins. They concluded that working with the sampler in *TinkerPlots 2.0* seemed to help students to recognise that the frequency of two heads and two tails approached 25% while the frequency of one head and one tail approached 50%. However, this experiment took place in only one school and on one school day.

It is commonly thought that students observe how data from an experiment converges on the theoretical distribution. In fact, Lee, Angotti and Tarr (2010), reporting on how 11–12 year olds used a computer simulation to decide which of six companies were producing fair dice, concluded that it was not the cycling between model and data that was critical but developing well-connected conceptual links between model and data. Konold, Madden, Pollatsek et al. (2011) suggested that constructing such a link was non-trivial for some students who appeared to lack a notion of a 'true' probability. Their subject appeared to distrust the idea that the theoretical probability was in fact the true probability exactly because the theoretical probability almost always failed to predict exactly what happened when the experiment was

repeated. Indeed, to them, it was the experimental probability that reported what really happened.

A teaching episode reported by Noll and Shaughnessy (2012) focussed on samples and sampling distributions in probability tasks. In this episode students were engaged in making inferences about both known and unknown mixtures of coloured objects (i.e. estimating population proportions) based on empirical data obtained from repeated sampling.

Researchers studied the impact of team teaching between the regular teachers and the investigators across six middle and high school classrooms. They concluded that teaching which focused explicitly on distributions, especially sample-to-sample variability, enhanced students' reasoning about empirical sampling distributions.

6.4.3 Understanding Sample Space

Bryant and Nunes (2012) conducted a literature review for the Nuffield Foundation on children's understanding of probability. They regarded working out the sample space as one of four key demands in learning about probability. Moreover, generating representations, such as tree diagrams, organised lists, and dot plots, based on sample space outcomes can support drawing conclusions and provide evidence for predictions (Fielding-Wells 2015, Kazak and Pratt 2015). In their earlier review of student learning of probability, Jones et al. (2007) also noted the importance of sample space but they reported a range of difficulties in a concept that was not as straight forward as might be thought. They quoted research that identified difficulties: 1. in identifying possible outcomes even in simple random experiments; 2. in systematically generating all outcomes; and 3. through failing to consider the sample space when determining probabilities.

Nunes, Bryant, Evans, Gottardis and Terlektsi (2014) reported on how to support generating and using the sample space in quantifying the probability of an event in primary grades. They

claimed that the conceptual schemas, such as classification, logical multiplication and ratio, which children begin to develop earlier in other domains (i.e. subtraction), can be used in understanding sample space. Nunes et al. designed an intervention study to test their conjecture that sample space could be taught in primary school by building on children's prior knowledge of these three concepts. In their study, one group of 10-11 year-olds participated in a teaching programme focussing on classification, logical multiplication and the use of ratios to quantify the probability of an event. Another group of participants (a comparison group) received instruction promoting mathematical problem solving that was not related to sample space and probability. The third group (a waiting-list control group) was taught by the class teacher and did not participate in a particular teaching programme until after the study. The study showed that the children in the intervention programme performed significantly better than their counterparts in both comparison groups. However, there was no significant difference between the problem-solving group and the unseen control group on any of the post-tests. According to Nunes et al., an instructional programme promoting the use of tree diagrams supported students' development of combinatorial understanding. This in turn was needed to understand how to generate a sample space by building on the concepts of classification and logical multiplication. After systematically identifying all possible outcomes and classifying those into favourable and unfavourable cases in the sample space, students used the ratios to quantify the probability of an event. We note however that the suggested approach in this intervention study is only applicable to limited situations where the classical probability definition is used, where the sample space is discrete rather than continuous, and where each possible outcome is equally likely.

The aggregation of cases (as favourable and unfavourable), mentioned by Nunes et al. (2014), is a crucial step in determining the probability of an event by using ratios. However studies by Francisco and Maher (2005) and Nilsson (2007) indicate that this idea was challenging to

students in complex probability situations. For example, Francisco and Maher's (2005) study showed that while students were able to list all possible outcomes in combinatoric problems, they had difficulties in identifying the sample space in a probability problem and, particularly, in determining the denominators of the probability ratio.

Nilsson (2007) focussed on the notion of sample space as a model for probability predictions in chance games. This study explored the strategies used by students (ages 12-13) when pairs were asked to distribute a set of markers on a game board numbered from 1 to 12 and to play the game against the other group by looking at the sum of two unusual dice. Students used the following pairs of designed dice in the game: (111 222) and (111 222), (222 444) and (333 555), (1111 22) and (1111 22), (2222 44) and (3333 55), where, for example, (111 222) represents a 6-sided die with three 1s and three 2s on its faces. In each of these four different game settings, an analysis of sample space for totals of two dice was required for making a decision about the distribution of markers on the game board. The study showed that students intuitively began to use what they considered as the sample space to decide the most/least likely totals in a given dice set-up. However, their focus was on the resulting sums by looking at only the proportions of numbers available on the individual dice, rather than examining the number of different combinations to get each sum. Hence, their incomplete sample space provided a limited model for their decisions in different dice set-ups.

Abrahamson (2009a, 2009b) reported on the single case of Li, an 11-year-old student, using a specially designed scoop, which collected four marbles from a large pot, containing green and blue marbles in equal numbers. Any one scoop therefore contained one of 16 equally-likely outcomes. First, Li was asked what would happen if the researcher were to scoop the marbles. Second, he was given card and crayons and asked to colour in all the different scoops. Third, Li was asked to create a combinations tower, in effect a histogram of the number of (say) green marbles in a scoop. These tasks lay a foundation for the Binomial

probability distribution, which is typically one of the first formal models used by statisticians and taught in an advanced statistics course at high school level and in an introductory statistics course at university level. For example, they are relevant to modelling one-dimensional random walk problems, especially for young students (e.g., Kazak 2010) and the distribution of gender in 12-children families (e.g., Biehler, Frischemeier and Podworny 2015).

In a detailed analysis of the clinical interview that took place around these three tasks, Abrahamson reported that Li's initial perception of the likelihood of events such as 2 green and 2 blue marbles was undermined by the need to construct the various permutations in the second and third tasks. Li saw no reason not to consider some of those permutations as redundant repeats. When the repeats were ignored, it seemed that there were five events (0, 1, 2, 3 and 4 blue marbles in a scoop), and there was no apparent reason for not thinking of these five as equally likely. According to Abrahamson, it was only when Li was able to make a 'semiotic leap' that he was able to use the tools to warrant his initial correct intuitive perceptions.

The use of such bridging tools might initially have been meaningless but, as the tools were well designed from a pedagogical and epistemological point of view, they led to semiotic leaps such as recognising why the events in the five-point sample space were not in fact equally likely. In the study reported by Pratt (2000), the students needed to re-align fairness away from the totals of two dice to the individual combinations in what Abrahamson would have termed a semiotic leap.

Given the difficulties students often encounter in generating and using sample space in probability contexts, Chernoff and Zazkis (2011) suggested a new term, 'sample set', as a bridging tool between student-generated lists of outcomes and the conventional sample space consisting of equiprobable outcomes. A sample set was used to refer to any set of all possible

outcomes of an event. For example, in Abrahamson's (2009a) four-marble task, {4 green and 0 blue, 3 green and 1 blue, 2 green and 2 blue, 1 green and 3 blue, 0 green and 4 blue} would be a sample set listing all possible outcomes of the scoop experiment. Unlike some students' thinking, this is not the sample space used in computing probabilities as ratios because the listed outcomes are not equiprobable. Consequently it leads to an incorrect answer as seen in Li's case (Abrahamson 2009a). Chernoff and Zazkis argued for a pedagogical approach that "without compromising mathematical rigour, acknowledges the learner and serves as a bridge between personal, sometimes naive, and conventional knowledge." (p. 19).

6.4.4 The Role of Modelling

For a typical statistician, a model can be imagined as a generator of data comprised of a main effect (signal) that explains much of the variation together with residual or unexplained variation, sometimes referred to as random error (Wild 2006). With modern software, computational models can actually generate data, akin to the statistician's way of thinking about the model. In Section 6.3.1, we discussed the differing epistemologies of probability. Depending on the given situation, probability can be interpreted as a theoretical solution based on an equiprobable sample space, a relative frequency in the long run, or a subjective degree of belief. Shaughnessy (1992) advocated a modelling perspective. As seen in several research studies in the Jones et al. (2007) review chapter, probability can be viewed as a tool for modelling uncertain situations and making simulation-based inferences (Watson, Jones and Pratt 2012).

Although several studies below have demonstrated some promise as to how a modelling approach might support aggregate thinking, in relating to distribution and sample space, learning to model is non-trivial. Indeed, speaking about science, Lehrer and Schauble (2010) emphasised the difficulties faced by novices. In fact, Pfannkuch and Ziedins (2014) proposed that more emphasis be placed on helping students to appreciate the purpose of modelling.

More specifically they suggested that models be categorised as ‘good’ or ‘bad’, or otherwise that no model currently exists. In that way, they suggested that students could engage in modelling activity either to use a good model, improve a bad model or create a model where one does not exist.

The Role of Modelling in Understanding Distribution.

Modelling promises to offer some leverage in dealing with the issues raised above about the challenge of connecting sample space, theoretical and empirical distribution. Konold, Harradine and Kazak (2007) used a data modelling approach in exploring middle school students’ understanding of distributions. The modelling activities in a series of tasks that focussed on a ‘data factory’ metaphor involved using *TinkerPlots 2.0* modelling capabilities to create a distribution that would match the expected data in the real-world, such as hair length of females and males. Using a similar approach, Lehrer, Kim and Schauble (2007) examined 5th-6th grade students’ use of *TinkerPlots 2.0* tools to model a distribution of repeated measurements of their teacher’s head. Student-generated models included an estimate value of the true length of the circumference using the median of the real measurements and the combination of some random errors, such as reading error and ruler error. Comparing simulation results in *TinkerPlots 2.0* with the actual data helped students revise their model. Both studies suggested these types of data modelling tasks with young students as a foundation for important ideas in statistical inference.

Prodromou and Pratt (2006, 2013) studied pairs of students aged between 14 and 16 years as they worked with a specially designed microworld where the students controlled the throw of a basketball. Control was exerted through sliders, which controlled variables such as the angle of release. These variables worked either deterministically or stochastically by changing the parameter value and varying the spread around that value, thus introducing variation into the basketball throw. Within this setting, Prodromou and Pratt (2006) focussed

on students' development of two perspectives on data generated by the computer simulations, which were called modelling and data-centric perspectives. They distinguished the two perspectives on distribution as they suggested different ways of perceiving variation. The researchers proposed that: 1. the modelling perspective emerged when students manipulated the tools controlling the position and spread of the distribution; 2. the data-centric perspective was revealed when students focussed their attention on variation and the shape of the emerging data. They also argued that being able to coordinate these two perspectives was essential in viewing data as a combination of signal and noise, which is a fundamental idea in statistical thinking (Konold and Pollatsek 2002).

Drawing upon the coordination of two perspectives on distribution, Prodromou's (2012) work with pre-service primary school teachers focussed on making connection between the empirical probability and the theoretical probability of the sum of two dice. The findings showed that pre-service teachers paid attention to the variation in the empirical data distribution (data-centric perspective on distribution) and the stability of the relative frequencies in the long run with a resemblance to the theoretical distribution (modelling perspective). A few of them also were able to make the connection from theoretical probabilities (modelling) to empirical probabilities (data-centric) as a way to make predictions.

The Role of Modelling in Understanding Sample Spaces.

Konold and Kazak (2008) highlighted the model fit idea to connect the empirical distribution and the expected (theoretical) distribution. Within this approach students tried to make sense of observed data with regard to a model when making a prediction; they sometimes revised their model on the basis of data. Students tended to make their initial predictions based on their experiences or beliefs about the likelihood of random events, which were often in conflict with the accepted theory. Konold and Kazak argued that engaging students in

developing the sample space in which the compound event occurred provided a theoretical model and facilitated their explanations for the distribution of actual and/or simulation data generated in *TinkerPlots 2.0*. They also suggested that by evaluating differences, or the fit, between the expected distribution based on the sample space and the distributions obtained from the simulations, they began to perceive observed data as a noisy version of the theoretical expectation (the signal) in relation to the size of data collected. Hence, this model fit approach provided a context to focus students' attention on sample space, which was often a challenging concept especially when students encountered compound events, as suggested by the studies mentioned in 6.4.3.

Most recently, the importance of the sample space analysis is also shown by the studies presented at the SRTL9, which investigated the role of building models in developing students' informal inference skills in games of chance (Fielding-Wells 2015, Kazak and Pratt 2015). In the context of a chance game seen on a popular television game show, Fielding-Wells (2015) discussed that structuring the sample space using a tree diagram provided a theoretical model and helped children (aged 10-11) make informal inferences based on the fit between the model and the data from experiments with the game device. In the context of another chance game involving the sum of two dice, Kazak and Pratt (2015) working with pre-service middle school mathematics teachers also reported on a case in which the probability model based on sample space emerged from engaging in both the combinatorial analysis of possible outcomes and empirical data both from playing the game physically and from simulations in *TinkerPlots 2.0*.

6.4.5 Discussion

Our review of research in this section suggests modelling as an emerging perspective for engaging students in probability contexts. This area of research is relatively new and still

exploratory in the sense that conjectures are still being formed about how to support students' understanding of probability and ideas using a modelling approach.

As seen in the studies above, one advantage of the modelling perspective is that it brings statistical and probabilistic ideas together. These examples generally involve focussing on the match between the data generated empirically and the expected distribution based on sample space. In several of these studies the role of technology is also worth noting in facilitating even very young students' understanding of probability. In addition, the modelling perspective appears to be relevant to promoting informal and formal statistical inference, which is addressed in Chapter 8 of this Handbook, while students are expected to draw data-based conclusions. Research specifically on modelling is reported in Chapter 7.

6.5 Conclusion

In this final section, we summarise in broad terms each of the three central themes. For more detailed findings of our analysis, please refer to the discussions in each of the three main sections. In addition to this broad summary, we consider gaps in the research and future directions.

This chapter has focussed on research into how students learn to address uncertainty and how teachers support them in that process. The focus has been on that type of uncertainty that is more or less quantifiable. That is to say we have not discussed research on somewhat less tangible aspects of uncertainty, such as the 'black swans' (Taleb, 2010), totally unpredictable events that can have dire consequences. While these other types of uncertainty are socially very important and interesting, the statistics educator is particularly concerned about situations that might incorporate randomness, quantified through probability. To this end, we have focussed here primarily on recent research, which we have contextualised within previous reviews of related research.

In the first section of this chapter, we discussed how the research on heuristics and biases has been re-presented as underpinned by dual process theory, potentially offering new insights into the many difficulties teachers and researchers have unearthed over the years regarding understanding probability. In particular, the new theoretical basis for the research on heuristics may point to innovative pedagogies to support the triggering of System 2 thinking when making judgements under uncertainty. We discussed some of the more promising research in this area. There needs to be further research to identify how, in Kahneman's terminology, System 2 might be better trained to recognise scenarios in which System 1's solution is likely to be biased. In Gigerenzer's terminology, research is needed to identify pedagogic approaches that lead to more accurate fast and frugal heuristics. The important theoretical distinction here is that Kahneman's ideas hold out little hope for improvement in System 1 but rather in identifying how better to use System 2, whereas Gigerenzer would focus on researching better heuristics.

In the second section, we elaborated further by considering the impact of how tasks are designed, technology is adopted, and more generally how students are taught on the development of probability as a concept. This research presents the clear conclusion that teachers are central if students are to develop the slow thinking of System 2 to manage in a more sophisticated way the quick intuitions of System 1.

The second section summarised how, post Jones et al. (2007), there has been an increasing number of research studies focussing on students' understanding of the relationship between experimental and theoretical probabilities with the availability of new technology tools. However, there is still a scarcity of research when the sample space is continuous and also in the area of subjective probability at the school level. We found no research on Bayesian methods at this level (see Chapter 15 for more on Bayesian methods). Pedagogical approaches, including task design, to bridging the three dominant interpretations of

probability need to be developed and tested in classroom settings. The second section ends by summarising what appear to be key aspects of how teachers might have a positive effect. Further research on task, tool and activity setting design is needed to identify how best to offer hands on purposeful experience that promotes discussion and prediction, and bridges different epistemological perspectives.

The third section points out that, perhaps driven by advances in the use of technology and in software development for educational purposes, probability can be presented as a mathematical model of (quantifiable) uncertainty. Indeed, such software allows the student to express their understanding of chance in the form of computational probabilistic models that can be executed. A modelling perspective on probability seems to offer a bridge that might help learners to coordinate the potentially confusing classicist, frequentist and subjectivist epistemologies of probability.

At the very least, when students create such models, they engage in activity that crosses any artificial boundaries that may otherwise have been set up between probability and statistics. Curricula have for many years tended to separate probability from statistics. Such a separation might render probability somewhat meaningless as students struggle to recognise any utility for the topic. Modelling approaches can counter that danger. As well as the examples described in the third section above, there are many others scattered in the book as a whole (see, for example, the chapter on informal statistical inference). Nevertheless, a modelling approach brings with it some new difficulties, touched on in the third section.

Most educational research on modelling in this field is recent because modern technological tools have opened up new possibilities; perhaps as a result, the promise that modelling offers to help learners link probability and statistics remains open to further exploration. There needs to be more exploratory research that clarifies how pedagogic approaches might exploit the potential of modelling for probabilistic learning, while providing pathways through the

obstacles for learners that no doubt will become more evident. One challenge is how to design tasks that make modelling seem purposeful to learners so that they can begin to engage with its utility or power. Another challenge is how to provide guidance on what makes a model effective. At the same time, there is still need for investigating the role of other visualisation tools (physical materials, diagrams and so on) and teacher scaffolding in promoting the modelling approach especially during off-computer tasks.

Although such research would be exploratory, there may be other research opportunities, which can test verifiable conjectures. Bryant and Nunes (2012) argue that much of the research on children's understanding of probability is based on good ideas but that its design is limited. They call for many more cross-sectional and longitudinal studies as well as intervention projects that test causal hypotheses about the factors involved in children's learning of probability. Testing causal hypotheses is difficult in educational research because there is an ethical dimension that resists the construction of randomised controlled trials. Nevertheless there are now some examples of where this has been possible and Bryant and Nunes call for more. The field is now relatively mature and this review alongside earlier ones may help to identify opportunities for this type of systematic research that tests well formulated hypotheses. Of course, there continues to be a need for exploratory studies in less well-developed topics, such as in the area of modelling, where clear and testable hypotheses are not yet available.

References

Abrahamson, D. (2009a). 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. (2009b). A student's synthesis of tacit and mathematical knowledge as a researcher's lens on bridging learning theory. In M. Borovcnik and R. Kapadia (Eds.), *Research and developments in probability education [Special Issue]. International Electronic Journal of Mathematics Education*, 4(3), 195-226.
- Abrahamson, D., Berland, M., Shapiro, B., Unterman, J., Wilensky, U. (2006). Leveraging epistemological diversity through computer-based argumentation in the domain of probability. *For the Learning of Mathematics*, 26(3), 39-45.
- 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, 55-80.
- Abrahamson, D., & Wilensky, U. (2007). Learning axes and bridging tools in a technology-based design for statistics. *International Journal of Computers and Mathematical Learning*, 12, 23-55.
- Ainley, J., Pratt, D. Hansen, A. (2006). Connecting engagement and focus in pedagogic task design. *British Educational Research Journal*, 32(1), 23-38.
- Babai, R., Brecher, T., Stavy, S, Tirosh, D. (2006). Intuitive interference in probabilistic reasoning. *International Journal of Science and Mathematics Education*, 4, 627-639.
- Bakhtin, M. M. (1986). *Speech genres and other late essays* (trans: McGee, V. W.). Austin, TX: University of Texas Press.
- Bakker, A., & Hoffman, 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.

- Bennett, D. (2014). Sticking to your guns: A flawed heuristic for probabilistic decision-making, In E.J. Chernoff, B. Sriraman (Eds.), *Probabilistic Thinking, Advances in Mathematics Education* (pp. 261-281). Dordrecht: Springer Science+BusinessMedia.
- Ben-Zvi, D., Aridor, K., Makar, K., & Bakker, A. (2012) Students' emergent articulations of uncertainty while making informal statistical inferences. *ZDM: The International Journal on Mathematics Education*, 44, 913–925.
- Biehler, R., Ben-Zvi, D., Bakker, A., & Makar, K. (2013). Technologies for enhancing statistical reasoning at the school level. In M. A. Clements, A. Bishop, C. Keitel-Kreidt and J. Kilpatrick (Eds), *Third International Handbook of Mathematics Education* (pp. 643-688). Springer Science+Business Media, New York.
- Biehler, R., Frischemeier, D., Podworny, S. (2015). Elementary preservice teachers' reasoning about modelling a “family factory” with TinkerPlots – A pilot study. *Proceedings of the International Collaboration for Research on Statistical Reasoning, Thinking and Learning, SRL-9* (pp. 146-175). Paderborn, Germany.
- Bodemer, N., Meder, B., Gigerenzer, G. (2014). Communicating relative risk changes with baseline risk: Presentation format and numeracy matter. *Medical Decision Making*, 34, 615-627.
- Borovcnik, M., & Peard, R. (1996). Probability. In A. Bishop, K. Clements, C. Keitel, J. Kilpatrick, and C. Laborde (Eds.), *International Handbook of Mathematics Education* (pp. 239-288). Dordrecht, The Netherlands: Kluwer.
- Bryant, P. & Nunes, T. (2012). *Children's Understanding of Probability: A Literature Review*. London: Nuffield Foundation.

- Budgett, S., Pfannkuch, M., Regan, M., Wild, C. J. (2013). Dynamic visualizations and the randomization test. *Technology Innovations in Statistics Education*, 7(2).
<http://escholarship.org/uc/item/9dg6h7wb>
- Canada, D. (2006). Elementary pre-service teachers' conceptions of variation in a probability context. *Statistics Education Research Journal*, 5(1), 36-64.
- Chaput, B., Girard, J. C., Henry, M. (2008). Modeling and simulations in statistics education, in C. Batanero, G. Burrill, C. Reading and 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.
- Chernoff, E. J. (2012). Recognizing revisitation of the representativeness heuristic: An analysis of answer key attributes. *ZDM Mathematics Education*, 44, 941-952.
- Chernoff, E. J., & Sriraman, B. (2014). *Probabilistic thinking: presenting plural perspectives. Advances in Mathematics Education*. New York, NY: 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.
- Chiesi, F., & Primi, C. (2009). Recency effects in primary-age children and college students. *International Electronic Journal of Mathematics Education*, 4(3), 259-274.
- Corter, J. E., & Zahner, D. (2007). Use of external visual representations in probability problem solving. *Statistics Education Research Journal*, 6, 22-50.
<http://www.stat.auckland.ac.nz/serj>
- Diaz, C., & Batanero, C. (2009). University students' knowledge and biases in conditional probability reasoning. *International Electronic Journal of Mathematics Education*, 4(3), 131-162.

- English, L. D., & Watson, J. M. (2016). Development of probabilistic understanding in fourth grade. *Journal for Research in Mathematics Education*, 47(1), 28-62.
- Fast, G. R. (2007). Analogical reconstruction of the probability knowledge of Zimbabwean female secondary school students, *Canadian Journal of Science, Mathematics and Technology Education*, 7(2-3), 149-181.
- Fielding-Wells, J. (2015). Young children's development and use of models as evidence in support of informal statistical inference. *Proceedings of the International Collaboration for Research on Statistical Reasoning, Thinking and Learning, SRTL-9* (pp. 98-106). Paderborn, Germany.
- Fischbein, E. (1975). *The Intuitive Sources of Probabilistic Thinking in Children*. Dordrecht, Holland: Reidel.
- Francisco, J. M., & Maher, C. A. (2005). Conditions for promoting reasoning in problem solving: Insights from a longitudinal study. *Journal of Mathematical Behavior*, 24, 361–372.
- Gal, I. (2005). Towards 'probability literacy' for all citizens. In G. Jones (ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 39-63). New York, NY: Springer.
- Garfield, J., delMas, R., & Zieffler, A. (2012). Developing statistical modelers and thinkers in an introductory, tertiary-level statistics course. *ZDM—The International Journal on Mathematics Education*, 44(7), 883-898.
- Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond heuristics and biases. In W. Stroebe and M. Hewstone (Eds), *European Review of Social Psychology*, Volume 2 (pp. 83-115). Chichester, UK: Wiley.

- Gigerenzer, G. (1993). The bounded rationality of probabilistic mental models. In K. I. Manktelow and D. E. Over (Eds.), *Rationality: Psychological and Philosophical Perspectives* (pp. 284-313). London: Routledge.
- Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In G. Wright and P. Ayton (Eds.), *Subjective Probability* (pp. 129-161). Chichester, UK: Wiley.
- Gigerenzer, G. (2012). Contribution to panel 'What Can Economists Know: Rethinking the Foundations of Economic Understanding' *The Institute for New Economic Thinking's (INET) Paradigm Lost Conference*, Berlin. Downloaded on June 5th, 2017 from <https://www.youtube.com/watch?v=DdEEwoKkfMA>
- Gigerenzer, G. (2014). *Risk Savvy: How To Make Good Decisions*. London, UK: Allen Lane.
- 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). *Exploring probability in school: Challenges for teaching and learning*. New York, NY: Springer.
- Jones, G. A., Langrall, C., Mooney, E. S. (2007). Research in probability: Responding to classroom realities. In F. Lester (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning* (pp. 909-955). Greenwich, CT: Information Age Publishing, Inc. and NCTM.
- 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.

- Kafoussi, S. (2004). Can kindergarten children be successfully involved in probabilistic tasks? *Statistics Education Research Journal*, 3(1), 29-39.
- Kahneman, D. (1996). On the reality of cognitive illusions. *Psychological Review*, 103(3), 582-591.
- Kahneman, D. (2011a). *Thinking Fast and Slow*. Farrar, NY: Straus and Giroux
- Kahneman, D. (2011b). Thinking Fast and Slow, Google Talks. Downloaded on June 5th, 2017 from <https://www.youtube.com/watch?v=CjVQJdIrDJ0>
- Kahneman, D., Slovic, P., Tversky, A. (1982). *Judgment under Uncertainty: Heuristics and Biases*. Cambridge, England: Cambridge University Press.
- Kapadia, R., & Borovcnik, M. (1991). *Chance Encounters: Probability in Education*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Kazak, S. (2010). Modeling random binomial rabbit hops. In R. Lesh, P. L. Galbraith, C. R. Haines, and A. Hurford (Eds.), *Modeling Students' Mathematical Modeling Competencies: ICTMA 13* (pp. 561-570). New York, NY.: Springer.
- Kazak, S. (2015). 'How confident are you?': Supporting Young Students' Reasoning about Uncertainty in Chance Games through Students' Talk and Computer Simulations. In A. Zieffler & E. Fry (Eds.), *Reasoning about Uncertainty: Learning and Teaching Informal Inferential Reasoning* (pp. 29-55). Minneapolis, MN: Catalyst Press.
- Kazak, S., & Pratt, D. (2015). Pre-service mathematics teachers' informal statistical inference when building probability models for a chance game. *Proceedings of the International Collaboration for Research on Statistical Reasoning, Thinking and Learning, SRTL-9* (pp. 138-145). Paderborn, Germany.

- Kazak, S., Wegerif, R., & Fujita, T. (2015a). Combining scaffolding for content and scaffolding for dialogue to support conceptual breakthroughs in understanding probability. *ZDM Mathematics Education*, 47(7), 1269–1283.
- Kazak, S., Wegerif, R., & Fujita, T. (2015b). The importance of dialogic processes to conceptual development in mathematics. *Educational Studies in Mathematics*, 90(2), 105–120.
- Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction*, 6, 59–98.
- Konold, C., Harradine, A., & Kazak, S. (2007). Understanding distributions by modeling them. *International Journal of Computers for Mathematical Learning*, 12, 217–230.
- Konold, C., & Kazak, S. (2008). Reconnecting data and chance. *Technology Innovations in Statistics Education*, 2(1), Article 1.
<http://repositories.cdlib.org/uclastat/cts/tise/vol2/iss1/art1>
- Konold, C., Madden, S., Pollatsek, A., Pfannkuch, M., Wild, C., Ziedins, I., et al. (2011). Conceptual challenges in coordinating theoretical and data-centered estimates of probability. *Mathematical Thinking and Learning*, 13(1-2), 68-86.
- Konold, C., & Miller, C. D. (2011). *TinkerPlots 2.0: Dynamic data exploration*. Emeryville, CA: Key Curriculum.
- Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education*, 33, 259-289.
- Konold, C., Pollatsek, A., Well, A., Lohmeier, J., Lipson, A. (1993). Inconsistencies in students' reasoning about probability. *Journal for Research in Mathematics Education*, 24, 393–414.

- Kustos, P., Zelkowski, J. (2013). Grade-continuum trajectories of four known probabilistic misconceptions: What are students' perceptions of self-efficacy in completing probability tasks? *Journal of Mathematical Behaviour*, 32, 508-526.
- Lecoutre, M. P., Rovira, K., Lecoutre, B., Poitevineau, J. (2006). People's intuitions about randomness and probability: An empirical approach. *Statistics Education Research Journal*, 5(1), 20-35.
- Lee, H. S., Angotti, R. L., Tarr, J. E. (2010). Making comparisons between observed data and expected outcomes: Students' informal hypothesis testing with probability simulation tools, *Statistics Education Research Journal*, 9(1), 68-96.
- 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), <http://escholarship.org/uc/item/1b54h9s9>.
- Lehrer, R., & Schauble, L. (2010). What kind of explanation is a model? In M.K. Stein and L. Kucan (Eds.), *Instructional Explanation in the Disciplines* (pp. 9-22). New York: Springer.
- Lehrer, R., Kim, M. J., 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. W. (2007). Teachers' understandings of probability. *Cognition and Instruction*, 25(2), 113-160.
- Meder, B. & Gigerenzer, G. (2014). Statistical thinking: No one left behind. In E.J. Chernoff, B. Sriraman (Eds.), *Probabilistic Thinking, Advances in Mathematics Education* (pp. 127-148). Dordrecht: Springer Science+BusinessMedia.
- Mercer, N., & Sams, C. (2006). Teaching children how to use language to solve maths problems. *Language and Education*, 20, 507-528.

- Nilsson, P. (2007). Different ways in which students handle chance encounters in the explorative setting of a dice game. *Educational Studies in Mathematics*, 66, 293–315.
- Noll, J., & Shaughnessy, M. (2012). Aspects of students' reasoning about variation in empirical sampling distributions. *Journal for Research in Mathematics Education*, 43(5), 509-556.
- Nunes, T., Bryant, P., Evans, D., Gottardis, L., & Terlektsi, M-E. (2014). The cognitive demands of understanding the sample space. *ZDM Mathematics Education*, 46, 437–448.
- Papariotodemou, E. (2011). Children's constructions of a sample space with respect to the Law of Large Numbers. In E.J. Chernoff & B. Sriraman (Eds.), *Probabilistic Thinking, Advances in Mathematics Education* (pp. 600-611). Dordrecht: Springer Science+BusinessMedia.
- Papariotodemou, E., Noss, R., Pratt, D. (2008). The interplay between fairness and randomness in a spatial computer game. *International Journal of Computers and Mathematical Learning*, 13, 89–110.
- Pfannkuch, M., Wild, C. J., & Parsonage, R. (2012). A conceptual pathway to confidence intervals. *ZDM – The International Journal on Mathematics Education*, 44(7), 899-911.
- Pfannkuch, M. & Ziedins, I. (2014). A modelling perspective on probability. In E.J. Chernoff, B. Sriraman (Eds.), *Probabilistic Thinking, Advances in Mathematics Education* (pp. 101-116). Dordrecht: Springer Science+BusinessMedia.
- Piaget, J., & Inhelder, B. (1951). *The Origin of the Idea of Chance in Children*. New York: Norton.
- Pratt, D. (2000). Making sense of the total of two dice. *Journal for Research in Mathematics Education*, 31(5), 602-625.

- Pratt, D., Ainley, J., Kent, P., Levinson, R., Yogui, C., & Kapadia, R. (2011). Role of context in risk-based reasoning. *Mathematical Thinking and Learning*, *13*(4), 322-345.
- 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.
- Pratt, D., Levinson, R., Kent, P., Yogui, C., & Kapadia, R. (2012). A pedagogic appraisal of the Priority Heuristic. *ZDM Mathematics Education*, *44*, 927–940.
- Pratt, D., & Noss, R. (2002). The micro-evolution of mathematical knowledge: The case of randomness. *Journal of the Learning Sciences*, *11*(4), 453–488.
- Pratt, D., & Noss, R. (2010). Designing for mathematical abstraction. *International Journal of Computers for Mathematical Learning*, *15*(2), 81-97.
- 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*, 126-154.
- Prodromou, T. (2012). Connecting experimental probability and theoretical probability, *ZDM Mathematics Education*, *44*, 855–868.
- Prodromou, T., & Pratt, D. (2006). The role of causality in the co-ordination of two perspectives on distribution within a virtual simulation. *Statistics Education Research Journal*, *5*(2), 69-88.
- Prodromou, T., & Pratt, D. (2013). Making sense of stochastic variation and causality in a virtual environment. *Technology, Knowledge and Learning*, *18*, 121–147.
- Ruthven, K., & Hofmann, R. (2013). Chance by design: devising an introductory probability module for implementation at scale in English early-secondary education. *ZDM Mathematics Education*, *45*, 409–423.

- Schnell, S., & Prediger, S. (2012). From “everything changes” to “for high numbers, it changes just a bit”. *ZDM Mathematics Education*, *44*, 825–840.
- Shaughnessy, M. (1992). Research in probability and statistics: Reflections and directions. In Douglas A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 465-494). Reston, VA: National Council of Teachers of Mathematics.
- Smith, T. M., & Hjalmarson, M. A. (2013). Eliciting and developing teachers' conceptions of random processes in a probability and statistics course. *Mathematical Thinking and Learning*, *15*(1), 58-82.
- Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioural and Brain Sciences*, *23*, 645-726.
- Taleb, N. N. (2010). *The Black Swan: The Impact of the Highly Improbable*. London: Penguin.
- Watson, A., Jones, K., Pratt, D. (2013). *Key Ideas in Teaching Mathematics: Research-based guidance for ages 9-19*. Oxford, UK: Oxford University Press.
- Watson, J., & Callingham, R. (2013). Likelihood and sample size: The understandings of students and their teachers. *Journal of Mathematical Behavior*, *32*, 660– 672.
- Wild, C. J. (2006). The concept of distribution. *Statistics Education Research Journal*, *5*(2), 10-26.
- Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry (with discussion). *International Statistical Review*, *67*, 223-265.
- Weisburg, H. I. (2014). *Willful Ignorance: The Mismeasure of Uncertainty*. Hoboken, NJ: Wiley.

Zahner, D., & Corter, J. E. (2010). The process of probability problem solving: use of external visual representations. *Mathematical Thinking and Learning*, 12, 177-204.