Probability Education

Madhav Kaushish

Department of Mathematics, University of Arizona

Table of Contents

Probability Education

Probability is a relatively new field of mathematics. However, its applications are vast, including games of chance, Quantum Physics, Artificial Intelligence, strategic decision making, along with many other aspects of everyday life. Humans are notoriously bad at Probability, especially intuitively, as demonstrated by mountains of research in Cognitive Psychology. This allows us to be exploited in Casinos, and leads us to bad decision making.

Borovcnik & Kapadia (2016) identify two reasons for teaching probability. The first is that probability affords a different, but equally important, way of thinking about the world than classical logic does – as Hacking & Hacking (1990) put it, probability gives us the tools to ‘tame chance’. The second is that probability allows application of mathematics to realistic situations.

In this paper, I will first give an idea of the history of Probability and of Probability Education. For both of these, I will be using the timelines mentioned in Borovcnik and Kapadia (2016). The history of Probability Education will also touch on various ideas about Probability Education in the literature. Next, I will ask the question: What is Probability? This will explore some major interpretations of probability. The following section will be concerned with the findings of Cognitive Psychology in regards to intuitive human understanding of Probability. The research demonstrates that Probability is not natural and many of our intuitions are suspect. Hence, it gives support to the need for Probability Education and research into it. I will then move on to a discussion about Probabilistic Thinking followed by an overview of research into Probability Education, which will contain implications for educational practice.

History of Probability

The earliest known probability calculation is contained in a medieval European poem from around 1250 titled De Vetula, though the combinatorial techniques used were known in India and the Arab world much earlier (Bellhouse, 2000). De Vetula contains an account of a calculation regarding the odds of certain numbers appearing on the toss of two die.

While mathematicians like Cardano worked on Probability in the intervening years (Tartaglia, Cardano and Ferrari, n.d.), it was the correspondence between Pascal and Fermat which led to the formulation of the fundamental laws of Probability (Pascal, n.d.). Pascal and Fermat, continuing Cardano's work, started by addressing two problems called the Gambler's Ruin, to do with a very specific game of chance, and the Problem of Points, which was more crucial in uncovering the rules of probability. The Problem of Points is to do with the division of a stake when a game of chance is prematurely abandoned. Over the next century, the field grew, with mathematicians like Huygens, de Moivre, and Bernoulli, contributing to it.

The Reverend Thomas Bayes made one of the most significant contributions to Probability (A History of Bayes' Theorem, n.d.). However, his unpublished work would have been lost if it wasn't for Richard Price who edited it and got it published. Bayes' theorem allows us to engage in inductive reasoning in a rigorous manner. Laplace expanded the scope of Probability beyond games of chance to many scientific and practical areas, extending Bayes' work which he appeared to be unaware of. He introduced many of the fundamental results of Statistics. In the 19th Century, Gauss extended this work further, including the introduction of the Normal Distribution.

Kologorov was the first we know of to axiomatize probability theory (Gerovitch, 2013). This was as late as the 1930s. Over the next few decades, Probability Theory was eventually incorporated into the more general Measure Theory. Measure Theory is not traditionally a part of K-12 education, and hence I will be ignoring it.

History of Probability Education

Like many other areas of Education, the history of Probability Education appears to have begun with Piaget. Piaget & Inhelder (1976) describes the three periods of learning about chance and randomness. The first period, from about age 4 to 7, a child cannot distinguish between necessary and possible events. The child oscillates between thinking of events as predictable or unpredictable, but in principle, nothing is predictable or unpredictable. This allows the very young to accept miracles as natural.

The second phase, from about age 7 to 11, involves the development of an unsophisticated version of the concept of random events. The child thinks of undetermined events as predictable as opposed to determined events which can be deduced by concrete logical operations. Random situations are categorized as such by antithesis to deductive necessity.

The third phase involves a more formal understanding of combinatorics and probability. Preadolescent children can now assess the total number of possibilities and the favorable possibilities. At this stage, they also develop the understanding of random distributions.

Unlike Piaget's theory which doesn't relate understanding to instruction, Fischbein (1975) sets out a direction for Probability learning which does. Fischbein separates primary intuitions, which are cognitive acquisitions are developed without the need for instruction, from secondary intuitions, which are formed by scientific education mainly in schools. Central to

Fischbein's ideas are the relationships between logical thinking, intuition and instruction (Greer, 2001).

Greer (2001) identifies three key principles for the design of effective probability instruction in Fischbein's work. The first is to build upon primary intuitions wherever possible and build secondary intuitions where necessary. Fischbein acknowledges that the primary intuitions may contradict normative probability rules. The second principle is the need for what Fischbein calls 'prefiguration of structures,' by which he means the use of representations as media by which to acquire abstract structures. The final principle is the need to deal with the child's cultural bias for deterministic thinking which has been acquired through exposure to various disciplines.

Kahneman and Tversky's work (2013) reinforced the ides that people already have intuitions about chance, many of which are undesirable. I will talk more about the work in this tradition later on in the paper.

One of the first major conferences on Statistics Education was ICOTS 1 in 1982. The group on Probability and Statistics was organized by Fischbein. Since then, as Borovcnik and Kapadia (2016) mention, there have been at least three major books compiling probability research: Chance Encounters (1991) edited by Kapadia & Borovcnik, Exploring Probability in School (2005) edited by Jones, and Probabilistic Thinking (2014) edited by Chernoff & Sriram.

What is Probability?

While I did go through an overview of the history of Probability in the Introduction, one issue was sidestepped: the different interpretations of Probability. This is an extremely important topic for Probability Education since these different interpretations have different consequences for how to think about Probability. There are various ways that the different interpretations have been classified. The simplest classification is into Probability as a feature of the world and Probability as a feature of knowledge about the world. However, for the purposes of this document, I will be going with the classification present in the Mathematics Education literature which divides the interpretations into three categories: classical, frequentist and subjective (Eichler & Vogel, 2013).

The classical interpretation of probability dates back at least to Laplace. Laplace defines probability as "the fraction whose numerator is the number of favorable cases, and whose denominator is the number of all cases possible" (Eichler & Vogel, 2014). This definition clearly requires all the cases we are dealing with to be equiprobable, and it lacks the ability to deal with non-finite cases.

The frequentist approach began with the advent of Statistics. The law of large numbers allows us to define probability as the number around with the relative frequency of an event fluctuates (Renyi, 1992, cited in Eichler & Vogel, 2014). The problem with this approach to probability is that it constrains us to situations where an experiment can be repeated a large number of times.

The subjective approach equate probability with degree of belief. This is closely connected with Bayes' Theorem, which allows us to update our degree of belief in a proposition or an event in the world when given evidence which supports or doesn't support it. This approach makes probability about our understanding of the world rather than a characterization of the world itself (Bantanero et al., 2005). The main problem many find with this approach is that it gives no direction in finding prior probabilities (Borovcnik & Kapadia, 2014).

The important thing no note is that these approaches are not logically contradictory – coming to that conclusion would be a category error since they are talking about different things (Borovcnik & Kapadia, 2014). Taking the example of a coin toss, the classical approach is saying that given certain assumptions, the probability of a head is ½. The frequentist approach would say that the probability of a head is about ½ because we have run a large number of trials of that event and a head occurs about half the time, assuming the coin is actually fair. The subjectivist could state any value between 0 and 1 as their prior belief in the coin coming up heads. This doesn't contradict the others since they are talking about their belief and not reality. They would then update their belief using Bayes' Theorem, and get closer and closer to ½ assuming the coin is actually fair. The subjective approach is what is used in applications like Artificial Intelligence and military modeling.

Probabilistic Biases

The Heuristics and Biases tradition in Cognitive Psychology, which traces its roots to work by Daniel Kahneman and Amos Tversky, contains a large amount of research on what it calls Probabilistic Biases (Gilovich, Griffin, and Kahneman, 2002). These are certain human tendencies which deviate from what we would normatively consider to be the right conclusion in the domain of Probability.

While there are many probabilistic biases, I will highlight a few of the more important ones here. The classification and basic definitions are from Benjamin (2018).

Bias about Base Rates

There is a tendency for humans to ignore base rates. Kahneman's iconic example about blue and green cars illustrates this quite well. The scenario goes as follows:

A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data:

- 85% of cabs in the city are Green and 15% are Blue
- A witness identified the cab as Blue. The court tested the reliability of the witness under the circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colors 80% of the time and failed 20% of the time.

What is the probability that the cab involved in the accident was Blue rather than Green? (Kahneman, 2011).

Once we work out the consequences using Bayes' Theorem, we see that the probability is 41%. However, the most common answer people give for this question is 80%. What they are doing is ignoring the base rate, the proportion of Blue and Green Cabs in the city.

As can be seen, this bias can have significant impacts on the legal process. It can have similar impacts in the field of medicine. The often cited example of this is from Eddy (1982) related to the probability of having breast cancer when a particular test for it is positive. The false negative rate and the true positive rates are given. The doctors who have been asked to work through this example have been wildly off in their conclusions.

Biases about Random Sequences

Such biases exist in two forms, the Gambler's Fallacy and the Hot Hand Fallacy, which work in opposite directions. The Gambler's Fallacy is the belief that if you have been losing so far, there is a higher chance that you may win in the next go even though the believer knows that the events are independent and identically distributed (i.i.d.). For instance, if they constantly bet on heads on a coin flip and lost, they would put a higher likelihood than 50% that the next toss would be a head. The Hot Hand Fallacy is that when you are winning a lot, there is a higher chance of this continuing. Once again, this is when the believer knows the events to be i.i.d. What both of these achieve is people taking risks and bets outside of what they would otherwise be comfortable with.

These fallacies have been observed in lab situations (Ayton & Fischer, 2004), and in casinos (Croson & Sundali, 2005) where they have a significant impact on house profits. They have also been observed in judicial decisions in refugee asylum court, reviews of loan applications, and umpire decisions in baseball (Benjamin, 2018; Chen, Moskowitz, and Shue 2016). As far back as 1814, Laplace notices that people place more bets on numbers when they have appeared before.

While these two biases seem to take us in opposite directions, Rabin & Vayanos (2010) argue that the Hot Hand Fallacy arises from the Gambler's Fallacy. The motivation behind their model is that i.i.d. situations appear to have more streaks than one would naively expect.

Biases about Sampling Distributions

People tend to give a higher probability to an event when it is described as a sequence of sub-events than as a single event. Benjamin (2018) refers to this bias as 'Partition Dependence.' Tversky & Koehler (1994) found evidence for this in a study with undergraduates about natural causes of death. When asked for the probability of death by particular natural causes, the sum of those probabilities was significantly more than the probability the students assigned to the whole class of 'death by natural cause.' The same was found with physicians and other experts, but the results were not as dramatic (Fox & Clemen, 2005).

The Law of Large Numbers is also something which appears to be non-intuitive. Kahneman (2011) gives the example of rural American counties having both the highest and lowest rates of kidney disease. When people hear one or the other, they come up with some story about why that is the case. However, they are both the consequence of rural counties having small populations and hence more variance. One explanation of this is that there is confusion between distributions within a sample and distributions of means across samples (Sedlmeier & Gigerenzer, 1997).

There are other biases related to Sampling Distributions such as the overweighting of the mean and the overweighting of tail events.

Biases related to Belief Updating

Neoclassical Economics assumes humans update their beliefs using Bayes' Theorem. It is not just that this is wrong, but it is systematically wrong. In general, people tend to underestimate probabilities when presented with new information. Kahneman & Tversky present the example of two large decks of cards, one with 2/3s marked 'X' and 1/3 of marked 'O', and the other with the frequencies flipped. When presented with a sample of 12 cards with 2/3 marked 'X' and 1/3 marked 'O', subjects said that the probability of it being drawn from the first deck is 70% when the correct posterior probability would be 94% (Benjamin, 2018). There was also very little change when the number of cards picked was increased from 12.

Other Probabilistic Biases

Neglect of Probability

Neglect of Probability, a term coined by Cass Sunstein (Kahneman, 2011), is when we either completely ignore small probabilities. Nassim Taleb brought attention to small probability but high impact events, which he calls Black Swan events (Taleb, 2007). The 2008 global financial crisis was one such example. Taleb contends that these 'fat-tailed' events are not treated seriously enough by policy makers and academics, resulting in events like global financial crisis.

On the other hand, we have the example of terrorism. When probability neglect is present here, people focus on the bad outcome rather than the likelihood of harm (Sunstein, 2003). Terrorists are able to make use of this to cause far more severe disruptions to society that would be warranted by their actions alone.

Conjunction Fallacy

This is when a something more specific is considered more probable than strictly more general. Kahneman & Tversky made up what they called the 'Linda problem.' They described Linda as follows:

> Linda is thirty-one years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in antinuclear demonstrations.

One of the questions they asked based on this description was: Which alternative is more probable?

- Linda is a bank teller.
- Linda is a bank teller and is more active in the feminist movement.

Apparently, about 85-90% of undergraduates at several major universities chose the second option (Kahneman, 2011).

Probabilistic Thinking, Reasoning and Logic

Borovnick & Peard (1996) see Probabilistic Thinking as similar to Geometric Thinking in that it consists of intuitive insight one gets from studying the subject, beyond a superficial knowledge of mathematical terms and procedures. Probabilistic Thinking is thinking in scenarios that allow for the exploration and evaluation of different possible outcomes in situations of uncertainty (Bantanero et al., 2016). The above review of various probabilistic biases, as well as the wide use of probability in various applications, shows us the importance of developing probabilistic thinking.

Reasoning is an epistemic mental activity (Shapiro & Kouri Kissel, 2018) which allows us to move from assumptions to conclusions. So, in order to think probabilistically, we are constrained by probabilistic reasoning. Falk & Konold (1992) conceptualize Probabilistic Reasoning as a mode of reasoning that refers to judgments and decision-making under uncertainty and is relevant to real life, for example, when evaluating risks (as cited in Bantanero et al., 2016). Under this conceptualization, it involves:

- Identifying random events in nature, technology, and society;
- Analyzing conditions of such events and derive appropriate modelling assumptions;
- Constructing mathematical models for stochastic situations and explore various scenarios and outcomes from these models; and
- Applying mathematical methods and procedures of probability and statistics.

In a similar vein, Borovcnik (2011), as described in Borovcnik and Kapadia (2016), says probabilistic thinking consists of:

- the ability to balance between psychological and formal elements when using a personal scale of probability;
- the understanding that there are no direct success criteria in random situations;
- the ability to discriminate randomness and causality; and
- the understanding of the difference between reflecting on a problem and making a decision.

Probability Logics, on the other hand, are formal systems which allow us to compute conclusions when given certain premises. There are a large number of Probability Logics in the philosophy literature including Propositional, First Order, and Modal Logics (Demey et al., 2018). This is a really interesting area of inquiry which could have implications for education. However, I have been unable to find much talk about Probability Logics in the education literature.

There is a sense in which Probabilistic Thinking is different to other types of mathematical thinking. As Borovnick & Peard (1996) put it, in the rest of mathematics education, you arrive at definite conclusions – statements are either true or false (ignoring 3-valued logic and Gödel's Incompleteness Theorems). There is transitivity – if A implies B and B implies C, then A implies C. This is not the case when making statements about events as we would in Probability. For instance, a set of non-transitive die (labeled D_1, D_2, … , D_n) have the feature that D_2 usually beats D_1, D_{i+1} usually beats D_i, and D_1 usually beat D_n (Grime, n.d.).

Added to that, a large number of the conclusions in Geometry, especially the Geometry taught in schools, do not go against our intuitions. In the domain of Probability, as demonstrated above, the conclusions can often times run against our intuitions. In Geometry, it may be the case that some aspects, such as proof, are non-intuitive. However, conclusions rarely are. The clearest examples of this are the child gender problem and the Monty Hall problem (Chernoff & Sriram, 2013, pp ix). The child gender problem goes as follows: somebody with two children tells you that one of their children is a girl. What is the probability that the other child is a girl as well? The correct answer is 1/3 since the space of possibilities is the set {BB, BG, GB, GG}. Since we are told one is a girl, we are reduced to {BG, GB, GG}. Hence, the probability of two girls is 1/3. However, most people respond with ½ (Chernoff & Sriram, 2013, pp x).

Probabilistic reasoning also requires a significant amount of understanding of a large number of concepts such as conditionals, risk, expectations, random variables, proportional reasoning, and so on (Bantanero et al., 2016). One of the major roles of probability education is to give students an understanding of such concepts.

Probability in various Curricula

While research into probability education is not entirely bound by national curricula/standards, it is definitely guided by them (Jones, 2007). I will go through a few here including the NCTM and Common Core standards, the NCERT (National Council for Education Research and Training, India), and the UK National Curriculum.

NCTM and The Common Core

For grades 3-5, the NCTM (NCTM, n.d.) lists the following as standards related to probability:

- describe events as likely or unlikely and discuss the degree of likelihood using such words as certain, equally likely, and impossible;
- predict the probability of outcomes of simple experiments and test the predictions;
- understand that the measure of the likelihood of an event can be represented by a number from 0 to 1.

For grades 6-8, they list the following:

- understand and use appropriate terminology to describe complementary and mutually exclusive events;
- use proportionality and a basic understanding of probability to make and test conjectures about the results of experiments and simulations;

- compute probabilities for simple compound events, using such methods as organized lists, tree diagrams, and area models.

From grades 6-8, the Common Core (Common Core State Standards Initiative, n.d.-a) has the following broad headings for standards related to probability and statistics:

- Develop understanding of statistical variability.
- Summarize and describe distributions.
- Use random sampling to draw inferences about a population.
- Draw informal comparative inferences about two populations.
- Investigate chance processes and develop, use, and evaluate probability models.
- Investigate patterns of association in bivariate data.

In High School, the NCTM (NCTM, n.d.) lists the following standards:

- understand the concepts of sample space and probability distribution and construct sample spaces and distributions in simple cases;
- use simulations to construct empirical probability distributions;
- compute and interpret the expected value of random variables in simple cases;
- understand the concepts of conditional probability and independent events;
- understand how to compute the probability of a compound event.

In the same period, the Common Core (Common Core State Standards Initiative, n.d.-b) has the following sub-headings for its standards:

- Interpreting Categorical and Quantitative Data
- Making Inferences and Justifying Conclusions
- Conditional Probability and the Rules of Probability
- Using Probability to Make Decisions

National Council for Education Research and Training (NCERT), India

Probability only gets a passing mention in the National Curriculum Framework Position Paper on Mathematics, 2005. However, from the NCERT textbooks, we can extrapolate the goals of the probability curriculum. The first time probability is dealt with explicitly is in the 9th grade. The 9th grade textbook chapter deals exclusively with experimental probability. The 10th grade chapter talks about theoretical vs experimental probability and talks about complementary events. The 11th grade talks about sample space, mutually exclusive events, combining probability, and introduces an axiomatic approach to Probability. The 12th grade chapter talks about conditional probability, combining probabilities, Bayes Theorem, Random Variables, and Bernoulli Trials. It also furthers the axiomatic approach to probability. The statistics chapter talks about distributions. The Indian curriculum is the only one I have so far encountered which works through axiomatic probability.

UK National Curriculum

During Key Stage 3 (Mathematics programmes of study: key stage 3, 2013), which refers to the 7th, 8th and 9th grades, the expectations from the students are the following:

- record, describe and analyze the frequency of outcomes of simple probability experiments involving randomness, fairness, equally and unequally likely outcomes, using appropriate language and the 0-1 probability scale
- understand that the probabilities of all possible outcomes sum to 1
- enumerate sets and unions/intersections of sets systematically, using tables, grids and Venn diagrams
- generate theoretical sample spaces for single and combined events with equally likely, mutually exclusive outcomes and use these to calculate theoretical probabilities.

In Key Stage 4 (Mathematics programmes of study: key stage 4, 2014), which includes the 10th and 11th grades, students are expected to:

- apply the property that the probabilities of an exhaustive set of mutually exclusive events sum to one
- use a probability model to predict the outcomes of future experiments; understand that empirical unbiased samples tend towards theoretical probability distributions, with increasing sample size
- calculate the probability of independent and dependent combined events, including using tree diagrams and other representations, and know the underlying assumptions
- calculate and interpret conditional probabilities through representation using expected frequencies with two-way tables, tree diagrams and Venn diagrams.

Teaching and Learning of Probability

Jones et al. (2007) found the following concepts in various national curricula: Chance and Randomness, Sample Space, Probability Measurement, and Probability Distributions. Looking at the small number I picked above, these still seem to be the valued concepts across these curricula. There are, of course, variations. For instance, the NCERT curriculum is the only one which includes the axiomatic theory of Probability. I will be using the four concepts identified by Jones as a way to classify the research in Probability. In each of the sections, if appropriate, I will start with Elementary School education, and move from there to Middle and High School.

Chance and Randomness

Chance refers to the lack of certainty, while randomness refers to phenomena which have short term unpredictability and long term stability (Gal, 2005; Jones, 2007). In this section, I will explore deterministic vs uncertain interpretations of events, interpretations of the language associated with chance, and random phenomena. Metz (1998) synthesized the research regarding the challenges these concepts pose for learners (as cited in Jones at al., 2007):

a) failure to interpret uncertainty in patterns that emerge over many repetitions of an event;
b) a belief that a person or device (e.g., spinner) can exert control over an event; and
c) a belief that some sort of order, purpose, or reason underlies events

Determinism.

It is clear from the research in cognitive psychology that even adults often fall prey to deterministic interpretations of events which are uncertain (Kahneman & Tversky, 2013) and young children are not immune from this (Fischbein, 1975). However, contradicting Piaget, there is research to show that students as young as four can identify situations which are indeterminate or unpredictable (Byrnes & Beilin, 1991; Horvath & Lehrer, 1998; Kuzmak & Gelman, 1986; Langrall & Mooney, 2005) even though they find it easier to identify deterministic situations than non-deterministic ones (Fay and Klahr, 1996; Langrall & Mooney, 2005).

In fact, as mentioned in an earlier section, it is often the case that the thinking required in other disciplines and especially in mathematics differs from the thinking required for probability (Fischbein & Schnarch, 1997; Langrall & Mooney, 2005). The understanding of probability, hence, requires instruction.

Language of Chance.

There are many different concepts which students assign to probability related words, many of which are not the conventional meanings (Fischbein, Nello & Marino, 1991; Watson &

Moritz, 2003; Jones et al., 2007). The word impossible is sometimes used to mean low probability, and certain is sometimes used to indicate that there is some chance (Fischbein, Nello & Marino, 1991). This was found with both elementary and middle school students. Amir & Williams (1999) found that one of the most common meanings of the word 'chance' in the population they studied was 'something which just happens.' A student gives the example of walking down a road and a stone on you from nowhere as an instance of chance. The use of the phrase 50-50 is also interesting. As Watson (2005) pointed out, students tend to use the phrase to refer to a situation with uncertainty, sometimes even when there are more than two options.

Konald et al. (1993) observed that the high school and college students he worked with used the phrase 'least likely' differently from the phrase 'most likely.' When asked which particular sequence of heads and tails was more likely, most students correctly said that they were equally likely. However, in the case of least likely, less than half the students said that was the case.

Randomness.

While a particular occurrence of an event may be uncertain, there is a lot we can say about distributions of such events. Children's understanding of randomness has been explored using tasks originally designed by Piaget & Inhelder. These include random mixture tasks, random distribution tasks and random draw – random generator tasks (Langrall & Mooney, 2005).

Random distribution tasks involve the construction of a distribution. They have provided very little insight into student understanding since all humans struggle with such tasks (Langrall & Mooney, 2005). As Shaughnessy (1992) put it, "human beings should never be responsible for trying the generate 'random' choices."

However, Pratt & Noss (2002, 2010) created a virtual environment where students fixed computer virtual simulations of random generators. They were designed so their configurations could be edited in such a way that they work properly. Initially, students were unable to distinguish the short term character of randomness from the long term aggregate character. However, after playing with the gadgets, they slowly appeared to achieve some understanding. The authors suggest the following heuristics while designing interventions:

- Enable the testing by children of their personal conjectures.
- Seek to enhance the explanatory power of knowledge that might offer a route to normalized knowledge.
- Construct a task design that will be likely to generate purposeful activity and tools that encourage the construction of utilities for the key mathematical concepts.
- Identify or design representations of key mathematical concepts that can be used as control points needed by the child to pursue their aim.

Random mixture tasks involve progressively mixing two or more different types of objects together. Piaget & Inhelder's (1976) original design involved a 'tilt' box, in which balls, initially separated by color, moved around as the box was tilted. They found that it was only after the age of 7 that students appeared to recognize that the balls were unlikely to return to their original positions and that the balls would mix. This was called into question by Paparistodemou, Noss, and Pratt (2002) who worked with children aged 6-8 via a computer game in which they were encouraged to create a random mixture of balls.

Random draw tasks involve picking objects from a collection while random generator tasks involve things like flipping coins and throwing die. Metz (1998) found that very young children do not recognize these tasks as involving chance. Even by the age of 8, though they do recognize many events as involving randomness, this is not universal. For instance, they do recognize a spinner as random but not picking marbles from an urn (Langrall & Mooney, 2005).

Sample Space

Hovrath & Lehrer (1998) suggest that sample space requires the combination of various cognitive skills. As they say:

> For example, consider the task of predicting the outcomes of spinning two spinners with three congruent regions labeled 1, 2, and 3 where the outcomes will be the sums of the two results (i.e., the numbers 2-6). First, one must recognize that there are different possible ways to achieve some of the outcomes (e.g., there are two ways to get a sum of 3, a 1 on the first spinner and a 2 on the second spinner and vice versa). Second, one must have a means of systematically and exhaustively generating those possibilities. Third, one must map the sample space onto the distribution of outcomes.

Piaget & Inhelder suggested that 7 year old children are able to list out the set of outcomes of a 1 dimensional experiment. However, this has been called into question (Jones, 1974, as cited in Jones, 2005). In place of Piaget's model, Jones et al. (1997) suggest an alternative based on levels:

> children reasoning at Level 1 do not identify all possible outcomes that could be randomly generated. Instead, they focus on what they believe is more likely to happen (or more deterministically, what they believe will happen). At Level 2, children consistently list all outcomes for one-dimensional experiments but are unable to identify all outcomes in two-dimensional situations. By Level 3, they have moved towards using a systematic strategy for listing outcomes in two-dimensional experiments and at Level 4 they have

adopted systematic strategies for generating all possible outcomes. (Langrall & Mooney, 2005)

There has been limited success in teaching students the concept of sample space. In a teaching experiment by Jones et al. (1999), even after instruction, some children still continued to be guided by subjective judgments about sample space. They suggested that this was due to the imposition of certainty on randomness.

Combinatorics and Sample Space.

The difficulty students have with exhaustively listing the sample space for compound events has been attributed to their combinatorial thinking abilities (Fischbein & Grossman, 1997; Jones et al., 2007). In a study with 14-15 year olds, Bantanero et al. (1997) classified the common errors made in combinatorial reasoning as:

- Misinterpretation of the problem statement – This can happen in various ways. For instance, they may convert a simple problem into a compound one. They may also have different definitions for words like interpreting distribute to mean share equally.
- Error of Order – This involves students considering order when it is irrelevant or not considering order when it is relevant.
- Error of Repetition – This involves students repeating individual instances when that isn't warranted or not repeating when it is.
- Confusing the type of Object – This involves considering identical objects to be different or different objects to be identical.
- Excluding some Elements – This involves leaving out elements which are important to the question.

- Non-systematic listing – This involves students using trial and error rather than a systematic, recursive procedure.
- Mistaken intuitive answer – Answers without justification.
- Incorrect arithmetic operations for finding the solutions – This involves using unrelated arithmetic such as addition or multiplication where it is not required.
- Not remembering the correct formula of a combinatorial operation that has been correctly identified
- Not remembering the meaning of the values of parameters in the combinatorial formula
- Faulty interpretation of the tree diagram

The instructional suggestion of the authors of this study is that instruction 'should also emphasize the translation of combinatorial problems into the different models, recursive reasoning and systematic listing procedures, instead of the mere centering on algorithmic aspects and on definitions of combinatorial operations.'

Modeling and Sample Space.

Konald & Kazak (2008) developed a modeling approach to dealing with probability, including the concept of sample space. In this approach, students construct a model to make sense of observed data. Initially, students tended to use their intuitions. However, as they compared more and more data to models, they slowly started to see real data as a noisy version of a theoretical model. This provided context to direct students' attention to sample space (Pratt & Kazak, 2018).

Working with preservice teachers, Kazak & Pratt (2017) used the 'river crossing game' to discuss the sample space of two dice. The game goes as follows: There are two distributions of counters on either side of a river as shown in Figure 1. Two players take turns rolling the dice. If they have a counter with the sum of the rolled dice, that counter crosses the river and is out of the game. The goal is to have no counter left on your side.

Initially, some of the teachers said that the blue side would win. They changed their minds on that as they played the game. This eventually resulted in them seeing the value in systematically listing out the elements of the sample space.

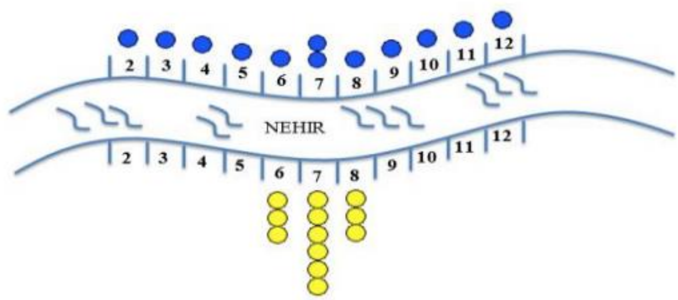

Figure 1. River Crossing Game

Probability Measurement

Classical Probability Measurements.

Classical probability tasks often involve students working through results of games of chance. An example given in Bantanero & Borovcnik (2016) is: In throwing 3 dice, is it better to bet that their sum is 9 or 10? What are the Probabilities? Acredolo et al. (1989, as cited in Jones et al., 2007) found the following strategies being adopted by students in such problems:

(a) a numerator strategy in which they only examined the part of the set that corresponded to the target event,

(b) an incomplete denominator strategy in which they examined the complement of the target event, and

(c) an integrating strategy in which they related the number of target elements to the total number of elements.

The importance of recognizing part-whole relationships as in c) above was shown by Jones, Langrall et al. (1999) to be crucial in conjunction with part-part relationships in order for 5th grade students to be able to understand quantitative probability.

A related type of task, called 'probability adjustment', was promoted by Falk & Wilkening (1998). They describe the task they use as follows:

> Participants are shown two urns with beads: one complete, with beads of both winning and losing color; the other incomplete, containing beads of only one color. Their task is to fill beads of the missing color into the second urn so that the probabilities of drawing a winning bead from either urn would be equal.

In their study, it was only by age 13 that students were using proportionality to answer this question, albeit at a less than perfect level (Jones et al., 2007).

Conditional Probability.

Conditional Probability is often confused with causality (Diaz et al., 2010). Students often believe that an event B cannot condition an event A if it happened after A. For example consider the following problem:

> An urn contains two white marbles and two red marbles. We pick up two marbles at random, one after the other without replacement.
>
> (a) What is the probability that the second marble is red, given that the first marble is also red?
>
> (b). What is the probability that the first marble is red, given that the second marble is also red? (Falk 1979, 1989 as cited in Diaz et al., 2010)

Students tend to accurately conclude that the answer to the first case is 1/3. However, in the second case, they conclude that the probability is ½.

Also, often, P(A|B) is confused with P(B|A). Bantanero et al. (1996) observed that many students confused “the percentage of smokers who get bronchial disease” with “the percentage of people with bronchial disease who smoke” (Diaz et al., 2010). Similarly, students tend to confuse mutual exclusiveness with independence (Diaz et al., 2010).

Diaz et al. (2010) compiled a list of various resources and tasks to teach conditional probability. The first type involve virtual environments where students can test their hypotheses about conditional probability. This includes simulations of the Monty Hall problem. Other resources include internet tools which help create tree diagrams, venn diagrams, and other representations of conditional probabilities.

Bayesian Probability.

Bayes’ Theorem emerges very easily as a result of the definition of conditional probability. However, the value of the theorem as a tool outside of mathematics is tremendous. An example of a basic problem involving Bayes’ theorem is the following:

> A person is indecisive about taking out a car insurance or not. We simplify matters and consider only two cases of damage: total wreckage (with a cost of € 30,000) and no accident. This person then has two possible decisions: d_1 taking the car insurance with a cost of € 1,000; and d_2 having no car insurance with costs depending on future outcomes.
>
> a. If the odds for total wreckage are 1:9, determine a value for each of the two decisions and find out, which decision is the better one.
>
> b. Find odds for total wreckage, for which the value of both decisions is the same so that both decisions are equally good. (Bantanero & Borovcnik, 2016)

If students are asked for their intuitive answer to questions like the one above, they usually get it wrong (Bantanero & Borovcnik, 2016). This was demonstrated in the section on cognitive biases.

There is another aspect to Bayes' theorem, Bayesian Updating. The goal of Bayesian Updating is to infer a degree of belief in a hypothesis when presented with new data. There have been criticisms of introducing Bayesian updating in an introductory Statistics course due to students having trouble with conditional probabilities, and the concept of priors (Albert, 2002; Moore, 1997). However, in some ways Bayesian Updating is far more intuitive than frequentist statistics (Albert, 2002), specifically that Bayes allows you to change your degree of belief about an event.

Probability Distributions

Scheaffer, Watkins, and Landwehr (1998, as cited in Pfannkuch & Reading, 2006) suggest that probability distributions ought to be the unifying thread throughout the probability curriculum. Probability Distributions, initially the normal distribution, are usually introduced to transition students from data analysis to statistical inference. Huck, Cross and Clark (1986, as cited in) identified two erroneous conceptions of standard scores. One group of students though z-scores only ranged from -3 to 3 since that is what they saw on most tables, while others thought that they have no limit. Bantanero et al. (2005) gave the following problem to students related to a computer file with data:

> In this data file find a variable that could be fitted by a normal distribution. Explain your reasons for selecting that variable and the procedure you have used.

The difficulties included:

> … perceiving the usefulness of theoretical models to describe empirical data; interpreting the probabilities under the normal curve; and discriminating between the cases where a discrete quantitative variable can and cannot be fitted by a normal distribution. Students also lacked the ability to use and interrelate all of the properties of the normal distribution when making a decision about whether the empirical distribution is approximately normal or not.

There has been some research what is called ‘distributional reasoning’. Variation is an important concept here. In a study with students from primary and secondary school, Reading & Shaughnessy (2006) found various ways students reason about variation:

- Concern with either middle or extreme values
- Concern with both middle and extreme values
- Discuss deviations from an anchor (not necessarily central)
- Discuss deviations from a central anchor

Their teaching recommendations based on this study were:

- We should not hesitate to give students more challenging tasks, given that they have access to computational devices
- The study of central tendency should not be separated from that of spread
- Students should be allowed untrained exploration of variation before more structured instruction
- Instructors should encourage students to justify their conclusions
- There should be a variety of tasks used

Conclusion

An understanding of probability is clearly valuable for various reasons. Unlike many other areas of mathematics, the ability to apply probability to real life situations is crucial for anybody living in today's world. However, the topic is non intuitive unlike many other areas of mathematics. In fact, in many aspects, it is counter-intuitive. On top of that, the field of Probability Education Research, like the field of Probability itself, is relatively new. Hence, we have a long way to go before we have a reasonable grasp on what probability education ought to look like.

It is also an open question as to whether and to what extent probability education can be successful in combating our biases. While it is unlikely to change our initial intuitions, what Kahneman calls our System 1, it could help us activate our System 2, the slow, deliberative part of our thinking, and use that in a rigorous manner when dealing with situations involving probability. In order to achieve our goals for probability education, we need psychologists, philosophers, scientists and mathematicians to work with education researchers in order to make further progress in this area.

References

A History of Bayes' Theorem. (n.d.). Retrieved from https://www.lesswrong.com/posts/RTt59BtFLqQbsSiqd/a-history-of-bayes-theorem.

Albert, J. (2002). Teaching introductory statistics from a Bayesian perspective. In Proceedings of the Sixth International Conference on Teaching Statistics. Cape Town, South Africa [CD-ROM]. Voorburg, The Netherlands: International Statistical Institute [Online: www. stat. auckland. ac. nz/~ iase/publications/1/3f1_albe. pdf].

Ayton, P., & Fischer, I. (2004). The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness?. *Memory & cognition*, *32*(8), 1369-1378.

Batanero, C., Chernoff, E. J., Engel, J., Lee, H. S., & Sánchez, E. (2016). Research on Teaching and Learning Probability. In C. Batanero, E. J. Chernoff, J. Engel, H. S. Lee, & E. Sánchez, *Research on Teaching and Learning Probability* (pp. 1–33). https://doi.org/10.1007/978-3-319-31625-3_1

Batanero, C., Henry, M., & Parzysz, B. (2005). The Nature of Chance and Probability. In G. A. Jones (Ed.), *Exploring Probability in School* (Vol. 40, pp. 15–37). https://doi.org/10.1007/0-387-24530-8_2

Batanero, C., Navarro-Pelayo, V., & Godino, J. D. (1997). Effect of the implicit combinatorial model on combinatorial reasoning in secondary school pupils. *Educational Studies in Mathematics*, *32*(2), 181-199.

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

Benjamin, D. J. (2018). *Errors in probabilistic reasoning and judgment biases* (No. w25200). National Bureau of Economic Research.

Borovcnik, M. (2011). Strengthening the Role of Probability Within Statistics Curricula. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education* (Vol. 14, pp. 71–83). https://doi.org/10.1007/978-94-007-1131-0_11

Borovcnik M., Kapadia R. (2014) A Historical and Philosophical Perspective on Probability. In: Chernoff E., Sriraman B. (eds) Probabilistic Thinking. Advances in Mathematics Education. Springer, Dordrecht

Borovcnik, M & Kapadia, R. (2016). Reasoning with Risk: a Survival Guide. 10.13140/RG.2.1.2077.0803.

Borovcnik M., Peard R. (1996) Probability. In: Bishop A.J., Clements K., Keitel C., Kilpatrick J., Laborde C. (eds) International Handbook of Mathematics Education. Kluwer International Handbooks of Education, vol 4. Springer, Dordrecht

Byrnes, J. P., & Beilin, H. (1991). The cognitive basis of uncertainty. *Human Development*, *34*(4), 189-203.

Chen, D. L., Moskowitz, T. J., & Shue, K. (2016). Decision making under the gambler's fallacy: Evidence from asylum judges, loan officers, and baseball umpires. *The Quarterly Journal of Economics*, *131*(3), 1181-1242.

Chernoff, E., & Sriraman, B. (2014). Probabilistic thinking. *AMC*, *10*, 12.

Common Core State Standards Initiative. (n.d.-a). Statistics & Probability. Retrieved from http://www.corestandards.org/Math/Content/SP/.

Common Core State Standards Initiative. (n.d.-b). High School Statistics & Probability. Retrieved from http://www.corestandards.org/Math/Content/HSS/introduction/.

Croson, R., & Sundali, J. (2005). The gambler's fallacy and the hot hand: Empirical data from casinos. *Journal of risk and uncertainty*, *30*(3), 195-209.

Demey, L., Kooi, B., & Sack, J. (2018). Logic and Probability. *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition). Edward N. Zalta (ed.). URL = <https://plato.stanford.edu/archives/sum2019/entries/logic-probability/>.

Eddy, D. M. (1982). Judgment under uncertainty: Heuristics and biases. *Judgment under uncertainty: Heuristics and biases*.

Eichler, A., & Vogel, M. (2014). Three approaches for modelling situations with randomness. In *Probabilistic thinking* (pp. 75-99). Springer, Dordrecht.

Falk, R., & Konold, C. (1992). The psychology of learning probability. *Statistics for the twenty-first century*, 151-164.

Fay, A. L., & Klahr, D. (1996). Knowing about guessing and guessing about knowing: Preschoolers' understanding of indeterminacy. *Child Development*, *67*(2), 689-716.

Fischbein, H. (1975). *The intuitive sources of probabilistic thinking in children* (Vol. 85). Springer Science & Business Media.

Fox, C. R., & Clemen, R. T. (2005). Subjective probability assessment in decision analysis: Partition dependence and bias toward the ignorance prior. *Management Science*, *51*(9), 1417-1432.

Gal, I. (2005). Towards" probability literacy" for all citizens: Building blocks and instructional dilemmas. In *Exploring probability in school* (pp. 39-63). Springer, Boston, MA.

Gerovitch, S. (2013, August 22). The Man Who Invented Modern Probability - Issue 4: The Unlikely. Retrieved from http://nautil.us/issue/4/the-unlikely/the-man-who-invented-modern-probability.

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge university press.

Greer, B. (2001). Understanding probabilistic thinking: The legacy of Efraim Fischbein. *Educational Studies in Mathematics*, *45*(1-3), 15-33.

Grime, James. (n.d.). "Non-Transitive Dice." Singing Banana. Accessed September 29, 2019. https://singingbanana.com/dice/article.htm.

Hacking, I., & Hacking, T. (1990). *The taming of chance* (Vol. 17). Cambridge University Press.

Horvath, J., & Lehrer, R. (1998). A model-based perspective on the development of children's understanding of chance and uncertainty. *Reflections on statistics: Agendas for learning, teaching, and assessment in K-12*, *121148*.

Jones, G. A. (Ed.). (2006). *Exploring probability in school: Challenges for teaching and learning* (Vol. 40). Springer Science & Business Media.

Jones, G. A., Langrall, C. W., & Mooney, E. S. (2007). Research in probability: Responding to classroom realities. *Second handbook of research on mathematics teaching and learning*, *2*, 909-955.

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*(2), 101-125.

Jones, G. A., Langrall, C. W., Thornton, C. A., & Mogill, A. T. (1999). Students' probabilistic thinking in instruction. *Journal for Research in Mathematics Education*, *30*(5), 487.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kahneman, D., & Tversky, A. (2013). Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I* (pp. 99-127).

Kapadia, R., & Borovcnik, M. (Eds.). (2012). *Chance encounters: Probability in education* (Vol. 12). Springer Science & Business Media.

Kazak, S., & Pratt, D. (2017). Pre-service mathematics teachers' use of probability models in making informal inferences about a chance game. *Statistics Education Research Journal*, *16*(2).

Konold, C., & Kazak, S. (2008). Reconnecting data and chance. *Technology innovations in statistics education*, *2*(1).

Kuzmak, S. D., & Gelman, R. (1986). Young children's understanding of random phenomena. *Child Development*, 559-566.

Langrall, C. W., & Mooney, E. S. (2005). Characteristics of elementary school students' probabilistic reasoning. In *Exploring probability in school* (pp. 95-119). Springer, Boston, MA.

Piaget, J., & Inhelder, B. (1976). The Origin of the Idea of Chance in Children.(P. Leake, H. Fishbein, & L. Leake, Trans.) New York. *NY: WW Norton & Company*.

Rabin, M., & Vayanos, D. (2010). The gambler's and hot-hand fallacies: Theory and applications. *The Review of Economic Studies*, *77*(2), 730-778.

Mathematics programmes of study: key stage 3. (2013). Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/239058/SECONDARY_national_curriculum_-_Mathematics.pdf.

Mathematics programmes of study: key stage 4. (2014). Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/331882/KS4_maths_PoS_FINAL_170714.pdf.

Metz, K. E. (1998). Emergent ideas of chance and probability in primary-grade children. *Reflections on statistics: Learning, teaching, and assessment in grades K-12*, 149-174.

Moore, D. S. (1997). Bayes for beginners? Some pedagogical questions. In Advances in statistical decision theory and applications (pp. 3-17). Birkhäuser Boston.

NCTM. (n.d.). Data Analysis and Probability. Retrieved from https://www.nctm.org/Standards-and-Positions/Principles-and-Standards/Data-Analysis-and-Probability/.

Paparistodemou, E., Noss, R., & Pratt, D. (2002). Exploring in sample space: Developing young children's knowledge of randomness. In *Proceedings of the Sixth International Conference on Teaching Statistics, Cape Town, South Africa. Voorburg, The Netherlands: International Statistics Institute*.

Pascal. (n.d.). Retrieved from https://www.storyofmathematics.com/17th_pascal.html.

Pfannkuch, M. & Reading, C. (2006). Reasoning about distribution: A complex process. *Statistics Education Research Journal*, *5*(2), 4-9.

Pratt D., Kazak S. (2018) Research on Uncertainty. In: Ben-Zvi D., Makar K., Garfield J. (eds) International Handbook of Research in Statistics Education. Springer International Handbooks of Education. Springer, Cham

Pratt, D., & Noss, R. (2002). The microevolution of mathematical knowledge: The case of randomness. *The 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.

Sedlmeier, P., & Gigerenzer, G. (1997). Intuitions about sample size: The empirical law of large numbers. *Journal of Behavioral Decision Making*, *10*(1), 33-51.

Shapiro, S. & Kouri Kissel, T. (2018). Classical Logic. *The Stanford Encyclopedia of Philosophy* (Spring 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2018/entries/logic-classical/>.

Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions.

Sriraman, B., & Chernoff, E. J. (Eds.). (2014). *Probabilistic Thinking: Presenting Plural Prespectives*. Springer.

Sunstein, C. R. (2003). Terrorism and probability neglect. *Journal of Risk and Uncertainty*, *26*(2-3), 121-136.

Taleb, N. N. (2007). *The black swan: The impact of the highly improbable* (Vol. 2). Random house.

Tartaglia, Cardano and Ferrari. (n.d.). Retrieved from https://www.storyofmathematics.com/16th_tartaglia.html.

Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological review*, *101*(4), 547.