Nuffield Foundation

Children's understanding of probability

A literature review (full report)

Peter Bryant and Terezinha Nunes
University of Oxford

www.nuffieldfoundation.org

Contents

Nuffield Foundation
28 Bedford Square
London WC1B 3JS
Telephone: +44 (0)20 7631 0566
www.nuffieldfoundation.org

The Nuffield Foundation is an endowed charitable trust that aims to improve social well-being in the widest sense. It funds research and innovation in education and social policy and also works to build capacity in education, science and social science research.

The Foundation has funded this literature review, but the views expressed are those of the authors and not necessarily those of the Foundation.

ISBN: 978-0-904956-86-3

Foreword

In 2009, the Nuffield Foundation published *Key understandings in mathematics learning,* a review of the research literature on how children learn mathematics. It has been widely read and has already had an impact on mathematics teaching and policy in several countries.

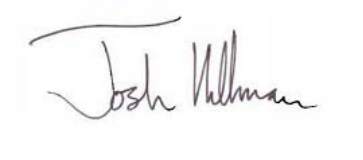
Probability was not included in *Key understandings,* and we subsequently commissioned two of its authors, Professors Peter Bryant and Terezinha Nunes from the University of Oxford, to examine the evidence on this topic.

There are four key reasons for our interest in probability. First, we wondered whether the teaching and learning of probability took sufficient account of children's prior knowledge of fairness, randomness and chance - concepts which are acquired at a very young age and which lay the foundations for probabilistic thinking. *Key understandings* noted that primary school geometry often failed to build on children's pre-school knowledge of spatial relations, and we thought probability might offer an interesting parallel. Second, the extent to which probability forms part of the primary curriculum has been subject to change in recent decades and so a consolidation of the evidence is timely. Third, this evidence is essential to underpin further research and development work, and fourth, probability is particularly relevant to our interest in statistical literacy in the wider population. Adults as well as children often find it difficult to think rationally about probability and randomness, so early encounters with these concepts are important.

In this review, the authors identify four 'cognitive demands' made on children when learning about probability, and examine evidence in each of these areas: randomness, the sample space, comparing and quantifying probabilities, and correlations. They draw together international evidence, from the early years through to adulthood, and highlight studies that are of particular relevance to teaching. They also identify areas that have been relatively neglected and would benefit from further research, particularly from fully evaluated intervention projects.

Indeed, the authors are currently seeking to address some of these gaps through a large-scale controlled study of the teaching of probability to 9-to-10-year-olds, which the Foundation is pleased to be funding.

We are grateful to the authors for their unstinting enthusiasm and commitment to this topic. The review is an informative and engaging read for anyone interested in how we understand (and misunderstand) probability, and provides valuable evidence that could be used to inform both teaching approaches and the design of future research.



Josh Hillman, Director of Education

Summary report

The ‘cognitive demands’ of understanding probability

Many of the events and relations in people’s lives are well understood and entirely predictable. If we knock a glass over, the liquid in it spills. If John is Michael’s father, John must be older than Michael. Other events and associations, such as a road accident or winning a lottery, are less predictable because they happen randomly. People know they might happen, but are uncertain if and when they will happen.

We can, nevertheless, reason logically about random events. This reasoning allows us to work out the probability of particular outcomes, and thus to understand the risks and possible benefits of acting in one way rather than another.

The understanding of the implications of randomness also lies at the centre of all statistical thinking. We decide the significance of any difference, for example in the recovery rates of patients given a specific drug and of others given a placebo, by calculating whether this difference could have happened by chance. Many associations, such as the association between income and health, are imperfect, and the most effective way of working out whether there is a genuine relation between two variables is to work out how much of the association could be due to random factors.

Randomness and uncertainty play an important part in scientific thinking as well, since many physical processes, such as the movement of subatomic particles are random, and need to be analysed in terms of probability.

Another good reason for people to be able to think rationally about randomness and uncertainty is that randomisation plays an important part in ensuring fairness in their every daily lives. Playing cards are shuffled and people are selected by lot to ensure that no one is given an unfair start.

Despite the central importance of randomness and probability in our lives, it is clear that children, and many adults as well, often have great difficulty in thinking rationally about, and quantifying, probability. Probability is quite a complex concept, and in order to learn about it we have to draw on our understanding of four different aspects of events and the sequence in which they occur. These four ‘cognitive demands’, as we call them in the report, are:

- ***Understanding randomness:*** To understand the nature and the consequences of randomness, and the use of randomness in our everyday lives.

- ***Working out the sample space:*** To recognise that the first and essential step in solving any probability problem is to work out all the possible events and sequences of events that could happen. The set of all the possible events is called ‘the sample space’ and working out the sample space is not just a necessary part of the calculation of the probabilities of particular event, but also an essential element in understanding the nature of probability.

- ***Comparing and quantifying probabilities:*** Probabilities are quantities based on proportions, and one has to calculate these proportions to make most (but not all) comparisons of the probabilities of two or more events. These proportions can be expressed as decimals, as fractions or as ratios.

- ***Understanding correlation (or relationships between events):*** An association between two kinds of event could happen randomly or, alternatively, could represent a genuine relationship. To discover whether there is a non-random relation or not, we have to attend to the relation between confirming and disconfirming evidence and check whether the frequency of confirming cases could have happened by chance. This means that, in order to understand correlations, we need to understand all three ideas mentioned above.

Randomness

Randomisation is a common and important part of everyday life, but it is clear that many adults' grasp of the nature of randomness and its consequences is quite tenuous. Research on young children suggests they have even more difficulty understanding randomness than adults.

Some aspects of randomness may be easier to understand than others. There are claims, for example, that even babies can understand the link between uncertainty and randomness. One study apparently shows that babies realise that choices made by people who cannot see what they are doing will be random, and governed by probability, whereas people who can see what they are doing will choose items that they want (Denison and Xu, 2009). However, problems with the design of this study mean it is not possible to reach a definite conclusion about this.

Piaget and Inhelder (1975) were the first to study children's understanding of randomness. In a classic experiment, they progressively randomised the position of marbles of two different colours, which were initially grouped by colour at one end of a tray, by tilting the tray and letting the marbles roll to the other side, and then by tilting it back and forth repeatedly. Young children could not predict the consequent jumbling of the two colours. However, this context was probably strange to the children, and the study needs to be done again with forms of randomisation, like shuffling cards, that are more familiar to children.

Research, using computer microworlds, has shown that by the age of about ten, many children realise that there is an association between randomness and fairness, and that randomisation can be an effective way of ensuring fair allocations (Pratt and Noss, 2002; Paparistodemou et al, 2008; Watson et al, in press). This association could be used to teach children more about the nature of randomness.

A common mistake made by adults and children, is to disregard the independence of successive events in a random situation. One's chance of getting a tail on the next toss of a coin is not affected by what happened on previous throws. Even if the last six throws were all tails, the result is no more or less likely to be a tail again on the next throw than it was on the first. Many people make the mistake of judging that, after a run of one kind of outcome, a different outcome is more likely the next time round. This is called the 'negative recency' effect. Another kind of mistake, called the 'positive recency' effect, is to predict after a run of one outcome that the same outcome is more likely to happen the next time. Many adults (Gilovich et

al, 1985) and most children make these mistakes, but recent research shows a higher proportion of positive recency errors among children than among adults and vice versa with negative recency errors (Chiesi and Primi, 2009).

Sample space

We can only calculate the probabilities of particular events if we know what all the possibilities are. The complete set of possibilities in a probability problem is called its 'sample space'. Working out the sample space is the essential first step in solving any probability problem (Keren, 1984; Chernoff, 2009), and in many it is the most important, since the solution is often quite obvious to someone who knows all the possibilities. Yet this aspect of probability has been relatively neglected in research on children's ideas about chance, which has concentrated for the most part on children's understanding of randomness and on their ability to quantify and compare probabilities.

Much of the information on people's awareness and use of sample space comes from mistakes that children and adults make in reasoning about probability, which they wouldn't have made if they had a thorough grasp of the relevant sample space (Fischbein and Gazit, 1984; LeCoutre and Durand, 1988; LeCoutre, 1992; Van Dooren et al, 2003).

In many probability problems it is necessary not only to list all the possibilities in the sample space, but also to classify them. This second step, which is usually referred to as `aggregation', can cause many children a great deal of difficulty. For example, if you throw two dice at the same time, there are 36 possible equiprobable outcomes (1,1; 1,2; 1,3 etc.). But, if you record the result in terms of the sum of the two numbers thrown, there are only 11 possible outcomes for the sums, which are two to12, and they are not equiprobable: a total of seven is twice as likely as a total of four, for example, because only three of the 36 possible pairs add up to four, whereas six of them add up to seven. Thus the individual outcomes are equiprobable but the aggregated outcomes are not. This difference causes great difficulty to some children (Abrahamson, 2009), and possible ways to address this would be an interesting question for further research.

The importance of the sample space also raises a general cognitive question, which is fairly obvious, but has never been discussed. To work out the sample space, the child must imagine the future in a particular way, and has to think of all the possible events that could occur in a particular context. There is some research on children's anticipation of particular and highly determined future events, but none on their ability to construct an exhaustive list of alternative, and uncertain, possibilities. Studies of this aspect of thinking about probability are sorely needed.

Quantifying probabilities

Probability is a quantity: it is a quantity based on proportions, and is usually expressed as a decimal number, a percentage or a ratio. The solution to most probability problems rests on the calculation of one or more proportions, but a few can be solved on the basis of simple relations like 'more' or 'larger'.

There is some evidence that even babies in their first year of life form expectations about the relative probability of two different possible events (Teglas et al, 2007; Xu and Garcia, 2008; Xu and Denison, 2009). They are surprised when someone draws mostly red balls from a container that they know to contain many more white than red balls. This reaction to an improbable outcome is evidence that they have some idea of the difference between probable and improbable outcomes. However, this is not evidence that they understand the proportional nature of probability.

Proportional reasoning in general, and not just proportional reasoning about probability, is difficult for young children. In the sphere of probability, this difficulty is most clearly illustrated by tasks in which children have to compare two or more different probabilities. Martignon and Krauss (2009) cite an example of this in a problem given to 15 year-olds: 'Box A contains one white and two black marbles. Box B contains two white and five black marbles. You have to draw a marble from one of the boxes with your eyes covered. From which box should you draw if you want a white marble?' The solution is not to be found in the absolute numbers of the two colours, but in the proportion of white marbles in each box. A large majority of the 15-year-olds given this problem made the wrong choice. Research by Piaget and Inhelder (1975), Falk et al (1980), Fischbein and Gazit (1984), and Falk and Wilkening (1998), does establish that pupils get better at making proportional calculations of probability as they grow older. However, there is no evidence to support Piaget's view that nearly everyone eventually becomes able to reason about probabilities proportionally. It is possible that many people never manage to do so effectively.

Proportions can be thought of, and calculated, in two ways. One is as a relation of a part to the whole. If a box contains two red and six blue marbles, the whole is all the eight marbles and the proportion of red marbles is 2/8 or 0.25, and of blue marbles 6/8 or 0.75, and this proportion is usually expressed as a fraction or a decimal number. The other is as the relation of one part to another, which is expressed as a ratio. In this example, the ratio of red to blue marbles is 2:6 or 1:3. There is good evidence that children come to understand proportions as ratios (part-part relations) before they understand them as fractions (part-whole relations) (Nunes and Bryant, 1996). This important distinction, however, has never been studied systematically in research on children's understanding of probability. Nonetheless, the reports of children's justification for their correct answers in probability comparisons in Piaget and Inhelder's (1975) and in Fischbein's research suggest that for the most part they used ratios rather than fractions in their reasoning (Fischbein, 1987; Fischbein and Gazit, 1984). The implication of this hypothesis is that children would learn about probabilities more easily if they are initially introduced as ratios.

In many instances, the probability of an event is dependent on the probability of another event. These conditional probabilities often cause adults, as well as children, a great deal of difficulty, as Kahneman and Tversky's (1972) work has established. An example of a conditional probability problem is a question about the likelihood that someone who has tested positive for a particular disease actually has that disease, when the incidence of the disease is 0.001 (or 1 in 1000) and the false positive rate for the test is 0.05 (or 5%). In this case the correct answer is dependent, not just on the false positive rate, but also on the incidence of the disease in the general population. Many people, however, attend only to the false positive rate of the test and not to the incidence of the disease, and this leads them to wildly incorrect calculations (in this example, to the incorrect answer that the probability is 0.95).

Recent research has shown that children and adults are much more likely to work out conditional probabilities correctly if the basic information is given as absolute numbers (one person in a thousand has the disease: five out of a hundred people who do not have the disease will test positive) than as decimal fractions (the probability of someone having the disease is 0.001, and the false positive rate is 0.05) (Hoffrage et al, 2002; Zhu and Gigerenzer, 2006). This interesting difference may be connected to the distinction between working with ratios and with fractions in probability problems. It is relatively easy to convert absolute numbers into ratios. Thus, the suggestion of teaching children about probabilities by first presenting these as ratios, rather than as fractions, may hold for conditional, as well as for simple, probabilities. It would be easy to do research on this idea.

Correlations

When two events happen together, their co-incidence might be either a random occurrence or the result of a genuine relationship. Since most such relationships are imperfect (taller people are usually heavier than shorter people but some short people weigh more than expected and some tall people weigh less than expected), we have to work out whether the imperfection of the association amounts to randomness or to a regular relation with exceptions. Thus, correlational thinking depends, at least partly, on an understanding of randomness.

Correlational thinking also depends on children realising that the way to work out whether an association is random or not is to consider the relative amount of confirming and disconfirming evidence. It is difficult to consider the relative amount of confirming and disconfirming evidence without systematic records and their quantification. When people use simple intuitive reasoning, they often fall prey to a confirmation bias: they pay more attention to the confirming than to the disconfirming evidence (Wason, 1968; Evans, 1989; Nickerson, 1998). Examples of this tendency are the idea that someone may have a winning streak in a casino, as if the turning of the roulette wheel had a connection with the player's choice, or that basketball players can have a hot hand, as if the fact that they scored in their last attempt makes it more likely that they will score again (Gilovich et al, 1985). Professionals working in clinical situations must be particularly aware of this confirmation bias: they see a biased sample of people and it is difficult for them to avoid this bias without systematic research (Chapman and Chapman, 1967; 1975). For example, if clinicians think that people only get better from problem *X* with a treatment that they prescribe, they must remember that the people who get better without the treatment are the people that they did not see, so they need to be aware of the risks of confirmation bias.

There is evidence that some adolescents do learn about the need to work out the relationship between the confirming and the disconfirming cases, and to do so proportionally (Inhelder and Piaget, 1958), but it is not clear yet how general this learning is. It is possible that only a minority learn to consider and relate the two kinds of evidence (Adi et al, 1978; Karplus et al, 1980; Batanero et al, 1996), and possibly only in situations where the two types of evidence can be systematically quantified and compared (Ross and Cousins, 1993). If this is the case, education should play a major role in people's understanding of correlation (Vass et al, 2000).

The future of research on children and probability

Research on children's understanding of probability has produced many interesting and educationally valuable conclusions, such as children's understanding of randomness in the context of fairness and the difficulties they have in reasoning proportionally in the context of probability. However, some aspects of children's reasoning about probability have been relatively neglected, such as the cognitive basis for constructing the problem space and the relative effectiveness of presenting and calculating proportions as ratios or as fractions. Another serious gap in research on children's ideas about probability is in longitudinal research, which is needed to establish how well children's early understanding and insights predict their overall learning later on, and also how complete their understanding of probability is by the time they leave school.

We make two main recommendations. The first is to take advantage of research designs that have been successful in research on other aspects of children's intellectual development. In particular, we recommend the combined use of intervention and longitudinal methods to study the links between the four aspects of probability, and to establish what experiences and abilities children need in order to learn about chance and uncertainty. This would provide a scientific basis for the effective teaching of probability.

Our second recommendation is that researchers on children's understanding of probability should pay much more attention to the great amount of related data that exists on other aspects of cognitive development. Probability makes a number of different cognitive demands and most of these demands are shared with other aspects of cognitive development about which we know a great deal. Probability is an intensive quantity, but so are density and temperature. Analyses of the sample space require combinatorial reasoning: so do many branches of scientific thinking. We think that many people doing research on probability have not paid attention to research on these related topics, and have missed out on potentially valuable information.

.....

Full report

Acknowledgements

We should like to thank Josh Hillman, Director of Education at the Nuffield Foundation, for his constant and enthusiastic interest in the report that we have prepared on children and probability, and for the good advice that he has given us. Dave Pratt was adviser to the project: his impressive knowledge of research on people's understanding of probability and the suggestions that he made for additions to the report were a great help to us. We also thank Tania Campos, Carmen Batanero and Jim Russell, who in different ways helped us to find out about aspects of the existing research on children's understanding of probability. Finally, we wish to express our deep gratitude to the Nuffield Foundation for making it possible for us to spend time thinking and writing about this exciting topic.

1. Introduction

What is special about probability?

Many of the things that we do and of the events that happen around us have consistent and entirely predictable outcomes. When it rains, the garden gets wet, and so do we if we go outside without an umbrella. We press a particular button and a radio starts, flick the right switch and a light comes on. Adults, and children too, readily understand how one thing can lead to another (Bullock and Gelman, 1979; Schultz, 1982), and they base a great deal of what they do and what they plan to do on straightforward causal sequences like these.

Other kinds of event and other actions have consequences that are not so certain, but are lawful nonetheless. We cannot say what will happen, heads or tails, when we toss a coin, but we do know a lot about what is likely to happen. If we toss it once, we are just as likely to throw a head as a tail, and if we toss it several times we are more likely to throw a mixture of heads and tails than just heads or just tails. The more times we toss the coin, the more equal will be the proportions of heads and tails in our throws. The sum of the two proportions will be 1, and so the probability of heads (expressed as a proportion) will be 1 <minus sign>- the probability of tails.

Our understanding of the probability of uncertain outcomes plays an extremely important part in our lives. We depend on it to decide about the medical treatment that we should follow, the insurance that we need, the car that we buy, and the precautions that we should take to protect our families and our homes. All these, and many other decisions depend on our knowledge of possible events that might happen and on our understanding of how likely these different events are. Sometimes we can use more information about events than just their general likelihood, and this informs decisions that we make as well as decisions that others make on our behalf. For example, if it is known that some areas in a city are more likely to be flooded than others, insurance companies may charge more for insuring homes in that area than in other areas; if we are buying a house, we may decide not to buy one in those neighbourhoods; and if flood prevention measures were taken to protect a neighbourhood, we might decide to risk it and buy a house there, although we know that preventative measures do not necessarily mean that there won't be a flood in the area. Similarly, if it is known that there is an association between being in a certain age range and being less likely to have a car accident, insurance companies may try to attract customers that cost them less money in the long run by offering them lower rates of insurance.

We also need this knowledge to deal with the ever-increasing use of statistics in modern life, since statistical comparisons are based on calculations about the probability of certain events, and can only be properly understood in that way. Our role as responsible citizens also depends increasingly on an understanding of probability, since many political and legal decisions are influenced by the assumptions about the probability and the improbability of certain events that are held by the people who make these decisions. Dawes (2001) suggests that in the last 100 years or so probabilistic thinking has been applied to everyday life, with serious consequences for citizens. 'Currently, for example, jurors are often asked to determine manufacturer or company liability on the basis of differential rates of accidents or cancers; in such judgments, jurors cannot rely on deterministic reasoning that would lead to a specification of exactly which negative outcome is due to what; nor can they rely on "experience" in these contexts in which they have had none.' (Dawes, 2001, p. 12082). So, jurors, who are faced with this kind of

problem, need to be able to reason, and to understand other people's reasoning, about probability.

These reasons for learning and knowing about probability may seem to have something of a negative tone, dwelling as they do on uncertainty, illness, floods, and accidents, but learning about probability has a positive and no less important side to it. One is that randomness is useful. It is an indispensable basis for being fair: shuffling cards to make sure that everyone has the same chance of getting a good hand, tossing a coin to decide who will start the serving or the batting, and distributing prizes and deciding which teams play each other in the opening games of the World Cup by lottery are clear examples of people using randomness to ensure fairness. Thus, randomness is an important part of everyday life. It is also, increasingly, an important part of science, quantum theory is about subatomic particles which move randomly: the central part of this theory is the Heisenberg uncertainty principle, and its implication is that we have to use the mathematics of chance to understand the physical movement of these particles.

We need, therefore, to understand probability in our everyday lives and as part of our intellectual understanding of the world around us. Children also need to understand probability. They, too, depend on randomness in formal and informal games, and they often have to deal with uncertainty. Yet, many people, adults as well as children, often find it hard to work out the probability of future events accurately, even in quite straightforward contexts and even though the calculations they need to make are often very simple indeed.

Four steps to understanding probability

Our starting point in this report on children's understanding and learning about probability is the sheer variety of the kinds of reasoning that are needed to solve problems about chance and uncertainty. This is illustrated quite dramatically by the many different kinds of mistakes that people make when reasoning about chance and uncertainty. Here are four examples of common, and yet diverse, obstacles to clear and logical reasoning about chance.

The first is the imperfect grasp that many people have of *the independence of individual events in a random sequence*. After tossing a coin and throwing three heads in a row, they judge that the next throw is much more likely as a result to be a tail than a head. This is a mistake, since a head is just as likely as a tail the next time, whatever happened in earlier throws, and it is a mistake with a name: it is usually called *the negative recency bias*. It has its mirror-image counterpart*, the positive recency bias,* which is the belief that an event that has happened by chance several times over is as a result more probable than before to happen the next time. Players, coaches and fans believe that, when a player scores in basketball, his chances of making the next scoring shot increase, although massive records of individual players in real games show this not to be the case (Gilovich, Vallone, and Tversky, 1985). Children, not surprisingly, also tend to make the same mistakes (Chiesi and Primi, 2009).

Our second example is about people *working out what the possibilities are*. This is an essential part of understanding probability, since in every probability problem there is bound to be a number of different possible outcomes. There are four possible outcomes (2^2) when you toss a coin twice (HH, HD, DH or DD), eight (2^3) when you toss it three times and so on. The number and the nature of all the possible outcomes is usually referred to as *'the sample space'*, and forming and then analysing this space is a crucial part of solving any probability problem, since

it determines how likely each outcome is. Yet, even well-educated adults often fail to imagine the form of the sample space correctly (Chernoff, 2009). In one study (Keren, 1984) the researcher told some university students a story about a game played by two boys (D and M) in which they jointly chose one card from a pack of cards. If the card was red, a coin was given to D and, if it was black, the coin was given to M. The boys did this three times and so the sample space for this story consisted of eight possible equiprobable events, which are (where D means D winning the coin, and M means M winning the coin) DDD, MMM, DDM, DMD, DMM, MMD, MDM, MDD. In only two of these outcomes is one boy the winner on all three occasions. In the remaining six cases, one boy gets two of the coins and the other one. Thus, the first of these two types of outcome – both boys win something, is three times as likely as the second – one boy scoops the lot. Yet 48% of the university students judged the two types of outcome as equally probable. They simply failed to work out the sample space.

Our third example is about *calculating probabilities*. Anyone who knows the sample space in a probability problem can then calculate a precise figure for the probability of particular events. This figure is usually a proportion, though it is sometimes given as a percentage or a fraction or a ratio. The proportional probability figure for each of the eight possible events in the problem that we have just described is 0.125; for the less probable of the two types of outcome, one boy scoops the lot, it is 0.25 and for the most probable of the two types of outcome, each boy gets something, it is 0.75. If the information were presented as a ratio, one would say that there is 1 chance in 4 that one boy scoops the lot and 3 chances in 4 that each of them gets something. One big advantage of the proportional figures is that they make it possible to compare the probabilities of different events with different sample spaces.

An example of a problem with this type of comparison comes from the 2003 PISA report (Pisa Consortium Deutschland, 2004, cited by Martignon and Krauss, 2009). This describes a problem given to a large number of 15-year-old German students: ‘Consider two boxes, A and B. Box A contains 3 marbles of which 1 is white and 2 are black. Box B contains 7 marbles of which 2 are white and 5 black. You have to draw a marble from one of the boxes with your eyes covered. From which box should you draw if you want a white marble?’ The absolute sample spaces are different in the two options but they can easily be compared by calculating the proportion of white marbles in A and in B: the answer is 0.33 in A and 0.29 in B, which makes Box A the correct answer. Yet, only 27% of these 15-year-old students (less even than chance-level) answered this question correctly, even though the mere choice of Box A with no further justification was counted as correct.

The fourth example is about *understanding the association between two different events*. In this case, the critical issue is to discriminate randomness from non-randomness. For some events, we can be relatively certain that one thing follows another: when we turn a particular knob, the radio goes on, or if we cut a finger, some blood will come out. But for many events in our lives, things are not so clear-cut. One example of strong, but by no means certain, associations is the diagnosis of illnesses. Illnesses are complex and the same illness may show a symptom in most patients but not in all. Conversely, the appearance of what may be a symptom of an illness in a person does not give us certainty that the person has this particular illness. So, many diagnostic tests have to be understood in terms of probabilities. Gigerenzer (2002) discusses the case of the association between the results of breast cancer screening using mammography and actually having breast cancer to illustrate this point. In women aged 40 to 50 who show no symptoms on the test but nevertheless have breast cancer, the probability of having a positive mammogram is 0.9. If a woman without symptoms and in the same age range does not have

breast cancer, the probability of her showing a positive mammogram is 0.07. Clearly, there is no certainty after a mammogram. Yet, mammograms are useful screening devices, and the association between the results of a mammogram and having cancer is not seen as fortuitous.

The association between two different events is technically known as a ‘correlation’. Understanding correlations involves understanding the three ideas discussed in the previous paragraphs: randomness, sample space, and the proportional quantification of probabilities. The key idea behind correlations is to test whether an apparent association between two events can be seen as fortuitous or whether it is unlikely to be fortuitous, given the calculations that we carry out about the probabilities of the two events occurring together. If we were assessing a cancer screening test and found that, of all the women who tested positive, 50% did have cancer and 50% did not, and the same turned out to be true of women who did not have cancer, we would have little doubt in saying that this was a useless test. But what if 70% of the women who tested positive had cancer and only 20% of those who did not have cancer tested positive? Would this be fortuitous or is the association stronger than that? Correlations help us to find an answer to this question.

We have chosen these particular examples not just to show that children and many adults find basic laws of probability difficult to use and even to understand, but also to make the point that learning about probability makes several different kinds of cognitive demands. Naturally, one has to know first what these demands are in order to work out what to do to help students overcome the difficulties that they entail. The framework that we adopt in this report will be that these demands fall into four categories, each of which corresponds to a basic and essential step that is present in all probability problems. Each of the four examples is an instance of one of these steps, which are:

1. Understanding the nature and the consequences of randomness
2. Forming and categorising the sample space
3. Quantifying probabilities
4. Reasoning about correlations.

One of the examples was about uncertainty due to randomness, another about the sample space, and the third about calculating probability precisely and the fourth about correlations and probability. In our framework, these correspond to three basic steps that one must always take in finding the solution to any probability problem, and a fourth step which is sometimes also necessary. The first step is to recognise that the problem is about outcomes that are uncertain because there is a random element in the frequency of their occurrence. The second step is to work out the sample space. In a probability problem the occurrence of a particular event is uncertain because there are other possible events and the probability of each event depends on what these alternatives are. Analysing the sample space solves this part of the problem. The third step is to calculate probabilities, and this consists of a proportional analysis of the sample space. The fourth step, which is not always needed, is too look for associations between variables in the samples space.

The intellectual demands of each of these steps are quite different from each other. They are also, for the most part, moves that children also have to learn to make in other contexts that have nothing directly to do with probability. For example, children have to reason about

proportions in scientific problems. Many basic scientific variables, such as temperature, speed and density are based on proportions. The density of an object, for example, is the ratio of its mass to its volume. Quantities such as density are called intensive quantities, in contrast to extensive quantities such as size and mass. The distinction between these two types of quantity is rarely, if ever, made explicit either in mathematics or in science lessons, but it is a crucial one in mathematical reasoning, because we operate with numbers differently depending on whether the quantities they represent are extensive or intensive. If you add one litre of water with a temperature of 20 ^{o}C to another litre with the same temperature, you double the volume of liquid (the extensive quantity) but its temperature (the intensive quantity) stays the same. Probability is also an intensive quantity. If you add one white ball and two blue balls to a container that already holds four white and eight blue balls, you increase the absolute amount of white and blue balls (extensive quantity), but the proportion of white and blue balls and thus the probability that you will pick a white ball from this container (intensive quantity) stays the same.

This is just one of very many examples of a cognitive demand that is an important part of learning about probability but is also at the centre of other kinds learning that children have to master at school. These underlying correspondences between different domains are immensely significant, because research on how children learn in other domains has produced a lot of information that is potentially of very great importance in learning about probability. Yet this potentially valuable link has often been ignored by people doing research on children's ideas about chance and uncertainty. It is one of the aims of our report to make this link and to show how useful it will be for future research.

How good is the evidence on children's learning about probability?

The report is about research that we think needs to be done as much as it is about research that has been done already. In reading about past research, we have, time and again, been struck by a clear contrast between the excellence of many of the ideas being tested and the ingenuity of many of the tasks used to do this testing on the one hand, and on the other hand the clear limitations of the design of the studies that contain these tests. Most research on children and probability takes the form of cross-sectional studies in which children of different ages are seen for a short period of time and are given tasks that test their understanding of one of the four aspects of probability outlined above. This kind of study tells us about what the children grasp, and what is difficult for them at particular ages, but it does not explain the reason why, as usually happens, older children do better than younger ones, and it give us no information about the connections between the four different aspects of learning about probability.

Research on other aspects of children's intellectual growth, such as learning to read and even learning about other aspects of mathematics, now rests on quite sophisticated research designs which do produce valuable information about the connections between different aspects of some kinds of cognitive growth. An example is the combined use of intervention and of longitudinal data to explore the connections between different aspects learning to read and write: this kind of research has shown quite clearly the importance both of knowledge and of morphological knowledge in children's literacy and has told us a lot about the interaction between the two (Nunes and Bryant, 2009). It seems to us that exactly the same kind of research could tell us about the possible causal connections, for example, between children's knowledge of the nature and the importance of the sample space and their ability to compare

the probabilities of a particular event in two different sample spaces. Probability needs this kind of research as much as literacy did.

One unfortunate result of the limitations in the design of research on children's understanding of uncertainty and chance is that the main hypotheses on the topic have never been properly tested. These hypotheses are about causal connections, which, as we have already remarked, are not properly tested by cross-sectional studies. Several prominent researchers, Fischbein (1987) among them, argue that even quite young children have some ideas about probability which are not appropriate ones, but which do provide a kind of platform for their learning more appropriate concepts. Fischbein claims that pre-school children have 'primary intuitions' about the nature of uncertain events, which they build for themselves on the basis of their own informal experiences. These primary intuitions are not coherent and they often lead to misconceptions on the part of the young child. However, they are the basis for the progress that children eventually make in understanding probability. With the help of teaching, they manage to reconstruct these initial ideas into much more successful 'secondary intuitions' about probability. Thus, two factors are responsible for the progress that children make, as they grow older, in understanding probability and solving probability problems. One is their initial primary intuitions and the other the experiences and the teaching that leads them to revamp these intuitions.

How does one test the hypothesis of these two causal factors? We will dwell here on the idea of combining longitudinal and intervention studies, which we have already mentioned in the context of research on children's literacy. One essential element of a proper test must be longitudinal research. If there really is a causal relationship, for example, between children's initial intuitions and the later growth in their understanding of probability, measures of one thing should predict measures of differences in the other in a longitudinal study. Thus, much of this longitudinal information should be about individual differences since, according to Fischbein's causal hypothesis, the children who develop these intuitions relatively early or relatively strongly should be better than others at solving probability problems months and even years later on.

Recently, as we will show in Sections 2 and 3 of this report, there have been some interesting attempts to establish the existence of some intuitive knowledge of probability in infants, but this will be very little use unless we can also establish some link between this early knowledge, if it really does exist, and the progress that the same children make in learning about probability later in their childhood. Unfortunately we have not yet been able to find any examples of longitudinal, predictive research in research on children's understanding of probability. This disappointing gap leaves us with no sure way of connecting the knowledge and the experiences of probability that children have early on in life and their eventual successes and failures in learning how to deal with chance and risk.

Another essential research method is intervention. Intervention studies play a dual role in research on children's intellectual development. One is to test hypotheses about what causes intellectual change in children. If you think, as many people do, that children's understanding of probability depends heavily on how well they can reason about and can calculate proportions, an excellent way to test this idea is to take steps to improve a group of children's proportional reasoning and then to see if this enhances their reasoning about chance and uncertainty as well.

The basic requirements for such a study are to assign children randomly either to an experimental group who, in this example, are given extra and effective teaching about how and when to calculate proportions or to a control group who are given just as much and as exciting extra teaching, but not about proportions, and to give all the children the tasks that are designed to measure their understanding of chance and probability before the extra teaching starts (pre-tests) and after it is over (post-tests). Since the effects of enhancing children's teaching in this way have often proved to be ephemeral and therefore not important, it is widely accepted that the children should be given post-tests not just immediately after the intervention stops but also after a delay of several months as well. The hypothesis about the importance of proportional reasoning would be strongly supported if this entirely imaginary study showed that the two groups had roughly equal scores in the pre-test, but the experimental group did much better than the control group in the immediate and delayed post-tests of their knowledge and understanding of chance. We are sad to admit that we have not found a single intervention study that has effectively measured the impact of proportional teaching on children's understanding of probability.

The hypothesis would seem even stronger if the positive result that we have just mentioned were combined with supporting longitudinal evidence that measures of children's proportional reasoning predict how well they learn about probability later on. Research on children's ideas about probability has included several intervention studies, which we will describe in the body of this report, but the approach of combining longitudinal methods and intervention, which has worked well in research on other branches of children's conceptual learning, such as literacy, has never been applied to probability.

We will end this discussion on the methods used to research this topic with the comment on the second role of intervention studies. They can measure the effectiveness of particular forms of teaching. The design of intervention studies designed to fulfil this aim should be much the same as the design of experiments designed to test causal hypotheses about the growth of ideas (pre- and post-tests and random assignment of the participants to experimental and control groups). In fact, it is quite possible that intervention studies could satisfy both aims at the same time. Properly designed intervention experiments, therefore, play a central role in theories about the causal connections in the growth of children's knowledge and understanding, and also in hypotheses about how to foster this growth at school and at home.

Summary

1. Children, as well as adults, need to know about uncertainty and chance. There is an obvious case for looking for effective ways of teaching children about probability.

2. The difficulties that adults and children have in reasoning about probability and solving probability problems are serious, but they are also diverse. They show that the understanding of chance makes a variety of cognitive demands.

3. There are four main aspects of successful reasoning about chance: (1) understanding randomness and its consequences, (2) analysing the sample space, (3) quantifying probabilities proportionally and (4) understanding and using correlations. Each of these makes different cognitive demands on children who are learning about probability.

4. Research on children's learning about probability has produced some interesting hypotheses and many ingenious tasks, but has not taken advantage of research designs that could establish the existence of the causal connections in these hypotheses. Coordinated longitudinal predictive studies and interventions could provide the data that we need for theories about the development of children's understanding and about how to encourage this understanding through education.

2. Randomness and its consequences

Uncertainty and randomness

Our starting point, whenever we try to work out the probability of some event, is uncertainty. We usually have a reasonable idea of the various events that could take place in the future, but we do not know exactly what will happen. Our uncertainty about the next event is due to the events themselves happening randomly. There is no discernible pattern, no set order, in the way that they occur and so there is no certainty about what the next event will be. Randomness is the hallmark of any probability problem. Mathematical analyses of probability are designed to deal with this uncertainty. If we have accurate information to hand about all the possible events, we can calculate the likelihood of each of them happening, even though they do take place in a random order, but we still remain unable to predict with any certainty which one will happen next.

At first sight, it seems likely that randomness should present no particular difficulty for children. Random sequences and random arrangements are a common part of their lives as well as of adults' lives. Many of their experiences are random. Balls bounce one way, and then another. Raindrops fall on one tile, but not on another. Randomness is also an essential part of some of the games that children and adults play or watch other people playing. Most sports games begin with a toss of a coin to decide who kicks the ball first, or serves first, or which side starts the batting. Throwing dice in Snakes and Ladders, in Ludo and in Monopoly is another overt form of randomisation. Its purpose is to ensure fairness: everyone has an equal chance of throwing the right numbers but we cannot say before the dice are cast who the lucky one will be. Card games have to begin with the shuffling of the cards, and shuffling is an obvious and very public way of ensuring a random sequence. It is an absolute requirement of all card games that no one should be able to tell which card will be dealt next, and the purpose of shuffling the cards is to produce this uncertainty. Again this tangible form of randomisation ensures fairness: everyone has an equal chance of being dealt a good hand.

Recognising randomness and distinguishing it from non-randomness

Although we may be confident that we know what random events are in everyday life, and can distinguish them from deterministic events, like lights switched on and off, some research suggests that even adults often cannot make this discrimination at all well. Our judgements about what are and what are not random sequences of events seem subject to biases. Falk and Konold (1997) illustrated one such bias by asking secondary school and college students in the USA to say whether certain patterns were random or not. The patterns were formed by rows of the letters X and O of different levels of complexity. The complexity of the row of letters was defined in terms of the numbers of chunks that would have to be memorized for the pattern to be reproduced. A row like XXXXXOOOXXXXXOOO, in which 5 Xs are followed by 3 Os, is of little complexity because it has basically two chunks repeated twice, whereas one like

XXOXOOOXXXXOOOXO, which has no easily recognisable pattern, has the same number of letters but more chunks to be memorised and is judged as more complex. Although all the sequences were random, adults tended to judge the relatively complex and less memorable sequences as random more often than the other sequences.

A second bias, identified by Tversky and Kahneman (1971, 1974) in research on psychologists, is the expectation that the characteristics of a large sample of random events will be replicated in a small sample of the same events. Tversky and Kahneman refer to this bias as a belief in the 'law of small numbers', a bias that leads us to expect, for example, that alternation between the letters in the previous example is more characteristic of a random sequence than a run of a few letters of the same kind. Dawes (2001) suggests that this is the reason for our tendency to judge patterns in which two possible outcomes alternate frequently as random and those with a run of the same outcome as determined. For example, if you toss a coin six times and obtain six heads, this result is judged as much less likely to have happened randomly than one in which a head is followed by two tails, followed by two heads and finally a tail. Yet, each of these sequences is equally likely and each has a probability 1/64. Dawes calculated that when people judge a sequence as truly 'random', they usually do so with a sequence that has a transition probability of 2/3 (i.e., different events follow each other 2 out of 3 times) rather than 1/2.

It is possible that judging outcomes as random or not 'intuitively' is rather difficult, but that we can learn to do so and be helped by cognitive processes other than perception. Given what we know about perceptual illusions, this is not at all unlikely, since we can override the perceptual effect of illusions with the help of reasoning and mathematical tools. For example, if we compare two circles of the same size, one surrounded by smaller circles and the other surrounded by larger circles, we see them as different: this is the well-known Titchener's illusion. However, we can measure both circles and conclude that they are of the same size, if their diameters are the same, even though we continue to see them as different (Rosengren and Hickling, 1994). Thus, the perception of randomness in a sequence and the cognitive understanding of random processes may not be the same thing. We should not assume a continuity between perception and cognition of randomness, as their differentiation may actually be at the heart of understanding probability.

Understanding the effects of randomisation

A large amount of research suggests that children also have a great deal of difficulty in making rational judgements about random sequences and the effects of randomisation. This research takes three forms: studies of children's understanding of the effects of randomisation, studies of their ability to discriminate random from non-random sequences and spatial patterns, and studies of their understanding of the independence of successive events in a random sequence. Many of the people doing research on children's understanding of randomness have looked at children's knowledge of randomisation. This seems a reasonable thing to do, since in many contexts randomness depends on effective randomisation. We can only be sure that cards will be dealt in a random sequence if they have been shuffled well or that lottery tickets will be picked in a random sequence if their container has been thoroughly shaken. These randomisations are such a commonplace in our lives that it is reasonable to expect that most adults will understand their nature and their consequences. It may not be the same with young children.

In one of the first studies on learning about chance, Piaget and Inhelder (1975) looked at the predictions that young children make about the consequences of randomisation. They showed the children, whose ages ranged from 4 to 12 years, some red and white beads, all neatly arranged in separate groups by their colour, and placed at one end of a tray. They asked the children to predict what would happen if the tray was tilted so that the beads all rolled to the other end of the tray, and then tilted again back to its former position, and then again. The most likely result, and therefore the answer that Piaget and Inhelder took to be correct, was that the red and white beads would mingle in an increasingly random arrangement each time that the tray was tilted. Most children over the age of about 11 years seemed to understand this.

However, many of the younger children, particularly the children in the 4- to 7-years range, predicted that the red and white beads would stay separate when the tray was tilted, either sticking always to the same side or swapping sides or both (i.e. swapping sides twice). Thus, these younger children apparently treated a chance event with several possible outcomes as a non-chance event with only one possible outcome. They usually agreed that tilting the tray would change the pattern, but they insisted that the new arrangement would be as ordered and as patterned as the initial one.

These results, according to Piaget and Inhelder, showed that younger children simply do not understand the effects of randomisation, and tended to treat a random spatial arrangement as a determined one. Their interesting argument was that children first have to understand deterministic cause-and-effect sequences before they can grasp the nature of random events. One of the main steps that they take in learning about deterministic cause and effect sequences is to realise that they are often ‘reversible’: a light can be switched off after it has been switched on and a car can be driven forwards and then backwards to its original starting point. Piaget and Inhelder claimed that children who have just learned about reversibility in causal chains tend to apply this idea quite inappropriately to randomisation as well, which is why they sometimes judged that all the balls of a particular colour would move sideways to the opposite side of the tray when it was tilted and then back again to their original position when it was tilted again.

However, some, at least, of the young children’s difficulties in this task might have been due to its unfamiliar physical context. The younger children may not have known much about what usually happens when several balls roll down a tilted surface. Their difficulty may have been in recognising that tilting a tray would randomise the arrangement of the two colours, and may have had nothing to do with their understanding the effects of a randomisation once it happens.

A study, by two American psychologists, Kuzmak and Gelman (1986), led these researchers to a conclusion that was quite different from Piaget and Inhelder’s. They presented children of 3 to 7 years with two different mechanisms, both of which dispensed balls of various colours one-by-one. In one mechanism the balls were lined up in an orderly way in a single tube, so that it was easy to see the colour of the ball that would come out next, while the other mechanism consisted of a complicated tangle of moving tubes which was designed to eject the balls in a random sequence and therefore made it impossible to predict the next ball out. The children were asked whether they could work out the colour of the next ball to be dispensed by each apparatus, and the majority of those who were 4- or more years-old did correctly say that they could make the prediction with the apparatus that was arranged in an orderly fashion but not with the disorderly apparatus, although that majority was very slim indeed in the case of the 4-year-olds.

Kuzmak and Gelman argued that this result showed that Piaget and Inhelder's contention about randomness was wrong and that even 4-year-old children can distinguish between random and non-random sequences. This conclusion was probably too hasty. The study did show that the children of 5 years or more generally made good judgements about their own knowledge of the next event: they really could say whether they did or did not know what that would be, but they could have made these judgements just on the basis of one arrangement – the tangled one – being a great deal more complicated than the other. They could easily have judged that it was harder to be sure what would happen with the complicated mechanism than with the relatively simple one, and thus their judgments may have nothing directly to do with randomness. This explanation would be in line with the finding by Falk and Konold (1997) that adults are more likely to judge as random those sequences of events that are too complex to memorise.

There is certainly room for a similar experiment with young children that uses other kinds of randomisation that are more familiar, more transparent and more comprehensible. Shuffling cards is an obvious candidate. One could start by showing each child a new pack of cards neatly separated into the four different suits and the cards in each suit arranged in exactly the same order as in the other suits, and then shuffle the pack very thoroughly. Both before and after this rearrangement the child could be dealt a card from a point somewhere in the pack and asked whether he or she could predict what the card next to it would be. The predictability in a shuffled deck of cards could be compared with that in a deck to which transformations were also made but these were not random, such as splitting the deck, moving the bottom half to the top, and then moving it once again to the bottom of the pile. As far as we know, there is no research of this sort, even though it is needed and would be easy to do.

Distinguishing random and non-random sequences

As we have seen, one way of looking at people's understanding of randomness is to find out whether they can distinguish a random from a non-random sequence. There is a long tradition of research on this question with adults, briefly referred to in the introduction. Adults turn out to be remarkably poor at distinguishing sequences that are completely random from others that contain a random element but are not fully randomised (Falk and Konold, 1997). The commonest mistake made by adults in this kind of task is to reject genuinely random sequences that contain quite long runs of the same value, like five successive heads in a row. It seems that many adults either do not know, or forget, that runs and patterns like this are perfectly possible and are to be expected in entirely random sequences.

This expectation of irregularity in a random sequence leads many adults into another confusion, which is to judge that irregular sequences with no particular pattern to them are much more likely in random situations than in regular ones. If I throw a coin six times, there are 64 (2^6) possible sequences that I could produce and they are equiprobable. One of these is a regularly alternating sequence starting with a head and ending with a tail – HTHTHT. Another is a run of six heads – HHHHHH. A third is an unbalanced and apparently irregular mixture – HTHTTT. Each of these sequences is as likely as the other and yet many adults wrongly judge sequences with no particular pattern as more probable than the regularly alternating ones (Kahenman and Tversky, 1972) in random contexts.

Children seem to behave to some extent like adults when judging the randomness of sequences. In a large-scale study of the understanding of randomness, Green (1979) gave UK secondary school children 3 different pictures of a square roof with 16 tiles on which a few (16

in each picture) snowflakes had fallen, and asked them to pick the most likely one. In one picture, 1 snowflake had fallen on each tile, in another the snowflakes had fallen in a regular pattern on all 12 peripheral tiles but on none of the 4 inner ones. In a third picture, the flakes made an entirely irregular pattern on the tiles, some of which were untouched, while two or more flakes had fallen on other tiles. Green called the third of these patterns the 'random' one. The children chose the pattern that Green called random less often than the pattern with one flake per tile, although no particular pattern can be seen as more likely than any other, as argued by Falk (1991. Batanero and Serrano (1999), who replicated Green's study, also provide more information about children's justifications in this problem. Children who chose the pattern of one flake per tile justified their choices in terms of the equiprobability of the outcomes. Thus they seem to be using, as adults do, the 'Law of Small Samples', and expected that what is observed in a large sample will also be observed in small samples. But their justifications also suggested that they did not see the sequence of events as independent: the distribution of one flake per tile was the result of this lack of independence of successive events.

Piaget and Inhelder (1975) tackled the question of children's understanding of randomness in a different way. They played a trick on the children by surreptitiously changing an entirely random mechanism into a non-random one that always produced the same outcome. The mechanism that they started with was a disc with a pointer, which the experimenter or the child could spin. This was at the centre of (and slightly above) a circular surface which was divided into equal size segments (like slices of cake) each with its own distinct colour: on top of each of these segments lay a distinctive box and these boxes could be, and were, shifted from one segment to another during the experiment.

At the start of the session, the spinner was not constrained in any way and so when it stopped the pointer might be pointing at any of the segments and thus at any of the boxes. Its position was quite random. After several spins, which produced a random sequence of results (sometimes the pointer stopped at the red segment, sometimes at the green and so on), the experimenters made a change without telling the children. They placed magnets in two of the boxes, which had the effect of stopping the pointer at particular points. On the whole, children below the age of roughly 7 years either did not notice the difference or were not greatly surprised by it, whereas older children were surprised and sought an explanation for these non-random sequences.

Piaget and Inhelder's conclusion that the younger children did not distinguish between chance and non-chance events deserves attention, but it may go too far. Runs of the same event, as we have already noted, do occur even in entirely random sequences, and we cannot be sure that the children didn't know this. We can wonder too about the trickery involved. The children may have been unwilling to change their judgements in the face of the new sequences because the experimenter, whom they trusted to explain what was going on, had said nothing.

Other researchers have examined children's reaction to biased, and therefore non-random, generators of outcomes that are usually purely a matter of chance. In a study, in which the ages of the children taking part ranged from 8 to 14 years at the beginning of the project, Watson and Moritz (2003) looked at these pupils' judgements about the 'fairness' of dice, by which they meant the equiprobability of throwing each of the six numbers on the dice. They gave the children three different dice, one of which was loaded: it was heavier on one side than on the others and as a result 5 was more likely than the other numbers to be the outcome of a throw. At the start of the project, 87% of the children either maintained that throwing die does not

produce equiprobable outcomes, or that it does, without any reason for either view: they appeared to the researchers to be making an arbitrary proposition. Nine percent of the children of the children either stated that the equiprobability of the outcomes depended on how well dice are made, or on the physical conditions of the throw itself. The remaining few of the children – a very small minority – argued that sometimes with a small number of throws the outcomes could be quite unequal, but that with a larger number of throws the outcomes would be generally approximately equiprobable. This minority therefore appealed quite correctly to the law of large numbers.

Is there an improvement over time in the quality of the children's beliefs? This was a partial longitudinal study: Watson and Moritz revisited rather less than half the sample four years later, and asked them the same questions about dice. Many of the children's beliefs had become more sophisticated over time, and yet the majority still expressed apparently arbitrary opinions about the fairness or unfairness of dice.

The picture that Watson and Moritz paint of the children's ability to reason about randomness is much bleaker than Piaget and Inhelder's, and it is clear that we need a lot more evidence on this issue. One useful question that this study throws up is about individual differences. It would be interesting, and extremely useful to know why some children's beliefs changed and improved over time while others kept to the same arbitrary assumptions over the four years of the project.

Recognising the independence of successive events in a random sequence

Each event in a random sequence is independent of any of the other events. What happened last time has no bearing on what happens next. The spin of a die is not in any way affected by the way that it spun last time which is why, having just thrown a 6 you are no more and no less likely to throw a 6 again on the following throw than you are to throw any of the other five numbers. If you have an urn full of an equal number of black and white balls and you pull out a succession of them, (taking care to replace each one after noting its colour) your chances of drawing black or white are just the same – 50% – every time. If, by chance, you start by pulling out five white balls in a row, the probability of your coming up with another white on your sixth draw is exactly what it was on the first draw – 50%. Of course, if you don't replace the balls each time, then pulling out five whites will actually increase the probability that you'll pull out a black one next time, but that is because you have radically changed the sample space, not because nature abhors a run of the same coloured balls.

The idea of the independence of random events is generally rather hard for humans to grasp. To go back to the example of the black and white balls, many adults would judge that a black ball is more likely to appear on the sixth draw after a run of five white ones than it was on the first draw. It is a very common mistake indeed, and it is a fundamental mistake because the independence of different events is an essential and fundamental aspect of random sequences. If what happens in the first five trials makes it possible for you to predict what will happen with the sixth draw better than you could predict the outcome of the first draw, the sequence is not a random one.

Yet, it is easy enough to see why even seasoned adults often make this error, which is usually called 'the negative recency effect' or, colloquially, 'the gambler's fallacy', and it is just as easy to sympathise with them for doing so. The fact is that, in the situation that we have described, it is much more likely that someone will pull out a sequence which consists of five white balls and

one black one in the first six draws than a sequence of six white balls. That is because there are six different possible sequences in which all but one of the balls is a white one (WWWWWB, WWWWBW, WWWBWW, WWBWWW, WBWWWW, BWWWWW) and only one sequence that is all white (WWWWWW). This means that sequences with one black and five white balls are six times more likely than an all-white sequence. The cause of the negative recency mistake is almost certainly people's knowledge that mixed sequences in general are more probable than unmixed ones in random situations. Thus, the mistake is probably a learned one. The people, who make it, could be misapplying the results of their experiences with random sequences in which they have seen many more mixed sequences than long runs.

Tversky and Kahneman (1974) have explained negative recency mistakes in this way: they attribute them to what they call the 'representativeness heuristic'. People see jumbled up and inconsistent sequences as a hallmark of randomness, and that makes them conclude that a black ball is more likely than a white draw after five successive white draws because randomness ensures inconsistency. Perhaps the best way to teach them that they are wrong in doing so is to show them that, though mixed sequences in general are more probable than runs, the two specific sequences WWWWWW and WWWWWB are equally likely, or in this case equally unlikely since each sequence has a probability of only 1/64 (p = 0.016).

There is another recency effect, called the 'the positive recency effect', which, as it name implies, takes the opposite form. Sometimes people take a long run of one outcome to mean that this is likely to continue: so, after choosing five white balls in succession, they think it more likely that their next draw will be another white ball than that it will be a black one. This idea fits well with folklore about people having runs of good (or bad) luck (it's their lucky/unlucky day). It is also a completely rational kind of judgement to make about non-random contexts (if flicking a light switch five times turns on a particular light each time, then it is very likely to have the same effect on the sixth, seventh and eighth occasions), but it is a bad basis for predictions about random events.

The existence of these two opposite mistakes, one probably based on what Kahneman and Tversky (1974) call 'representativeness', the other on folklore or on a confusion between random and non-random situations, raises a fascinating question about children's understanding of probability. If, as we have suggested, the negative recency mistake is the product of people's experiences with random sequences, one could quite reasonably expect that adults who have had more of this sort of experience than children have, would make the mistake more often than young children do. This would be a striking result since in very nearly every other cognitive task adults make fewer mistakes than young children do.

It has been known for some time that both adults and children make the two kinds of mistake, but the most systematic study of how common these mistakes are in the different age groups was only done quite recently. This is an experiment by Chiesi and Primi (2009) with 8- and 10-year-old Italian schoolchildren and a group of university students as well. They showed these participants pictures of bags containing balls of two different colours: sometimes the numbers of balls in the two different colours was equal, sometimes not, but for simplicity's sake we will begin by describing what the researchers did and what the results were when the frequency of the colours was equal.

The experimenters told the children the actual numbers of the two sets of balls – 15 green and 15 blue ones – and said that someone had already drawn 4 balls from the bag (replacing the

ball after each draw) and all 4 had been blue (or in other trials all green). This person was going to make another draw. The experimenters then asked for a prediction about that next draw: they gave the participants three choices:

1. It was more likely that the next draw would be blue ball than a green one
2. It was more likely that the next ball would be green than a blue one
3. The two colours were equally likely.

The third choice was the correct one. Picking the first choice represents positive recency and the second negative recency (when the first four draws had all produced a blue ball).

When the bag contained an equal number of balls, the members of all three age groups made a startlingly large number of mistakes. None of the 8-year-old children's answers was right: the 10-year-old children and also the college students were right only around 40% of the time. So, there was a marked improvement overall between the ages of 8 and 10years, but no evidence, from these overall scores, of any improvement after that.

What about the pattern of the mistakes made by each group? These could either be positive or negative recency mistakes and Chiesi and Primi found that there were striking differences with age in the proportion of these two kinds of error. There was a clear decline with age in the number of positive recency choices, and as clear a rise in the number of negative recency choices. This was both a relative and an absolute difference. The 8-year-olds made twice as many positive recency as negative recency choices; the 10-year-olds made roughly an equal number of positive and negative recency choices and the adults made nearly three times as many negative recency choices as positive recency choices. The negative recency bias in the adult group was so strong that they actually picked this option more times than either the 8- or the 10-year-olds did, even though the overall number of wrong choices that the adults made was a great deal less than that of the youngest group. So, the positive recency effect seems to decline as children grow older while the negative recency effect increases quite dramatically.

Understanding the independence of successive events from each other in a random sequence is a fundamental part of learning about randomness, and the clear implication of Chiesi and Primi's study is that there are surprisingly strong limitations to children's and even to well-educated young adults' grasp of this independence. However, the results of the study also suggest that the reasons for these limitations change over time. The very large number of positive recency choices among the youngest children (between 60% and 70% of all the choices that they made when the two colours were equal in number) suggests that their greatest difficulty with understanding the independence of successive random events may be due to a confusion between determined events, where the same action produces the same results time after time, and random sequences where outcomes of the same action cannot be predicted on each occasion. In contrast, the strikingly large number of negative recency choices made by the adults (43% of all the choices that they made when the two colours were equal in number) must be the result of a different confusion, and here Tversky and Kahneman's explanation in terms of representativeness seems the most plausible candidate.

Understanding the purpose of randomisation

Most definitions of randomness are couched in negative terms: they emphasise the uncertainty of the various outcomes in a random sequence or in a random spatial arrangement. But randomness has a positive aspect too, which is to provide us with a way of being fair. We have already noted that randomisation is an essential part of most games, either as a preliminary step, like selecting who bowls and who will be in Team A and who in Team B, or as a central part of the game itself like tossing dice in Monopoly and spinning a roulette wheel in a casino. In order to ensure that the randomisation is fair, it is usually open to everyone in the game. We all watch the captains tossing the coin at the beginning of the game and we pay great attention to the other Snakes and Ladders players when it is their turn to throw the dice. In card games, we even carefully monitor how thoroughly the cards are shuffled because we want them to be properly randomised so that we can be sure of a fair game.

The link between randomisation and fairness is, therefore, a frequent and very public part of games that most children play, and it poses obvious and hugely interesting questions for researchers. Do even young children understand the importance of shuffling cards and throwing dice in the games that they are taught to play? Does their experience of these forms of randomisation teach them anything about probability that they did not know already? How efficiently do children carry out these randomisation procedures themselves? Yet, a disappointingly few researchers have tried to answer these questions or to study children's reactions to randomisation procedures in games in any way at all.

The most notable exception to this bleak pattern is research by Pratt and Noss (2002) and Paparistodemou, Noss and Pratt (2008) on randomisation in a computer microworld. Pratt and Noss reported that the comments made by 10- and 11-year-old children working with random sequences showed that they started the study with some firm ideas about the nature of randomness. These ideas, which the researchers called 'inner resources', took the form of characteristics, which, the children thought, sequences must have in order to be random. There were four of these: the children thought of random sequences as 'unsteerable' (in other words, impossible to control), unpredictable, irregular and fair. These notions seem a good start to understanding randomness.

In the later study, Paparistodemou, Noss and Pratt (2008) introduced a group of 5- to 8-year-old children to a computer game that involved a prince who sat on a platform halfway between a blue (explosive) mine above him and a red mine below. There was also a little white ball that moved around the screen frequently bumping against and bouncing off some larger red and blue balls, which were static during the game. The path that the white ball took when it bounced off these other balls was random, but it determined the prince's position. Whenever it hit a blue ball, the prince moved up the screen towards the blue mine, and whenever it hit a red one the prince moved down towards the red mine. Thus, the effects of each kind of encounter cancelled out the effects of the other and restored the status quo and an imbalance of either blue or red encounters would mean the prince's obliteration. It was in the prince's interest that the white ball should have roughly the same number of encounters with blue and with red balls, and it was also in the interest of the children playing the game, since they were asked to ensure what the researchers called fairness, which meant an equal number of red and blue encounters, and thus to save the prince.

The children were allowed to change the position, the size and even the number of the static red balls before the game started, and the researchers wanted to know whether they would decide on a random arrangement of these static balls to ensure a kind of red–blue equality. With such an arrangement, the probabilities of the white ball hitting a blue ball and of it hitting a red one would be equal.

Most of the children did not adopt this strategy at the outset of the study. Instead, they preferred a symmetrical arrangement of the red and blue balls – red balls to one side of the screen, blue balls to the other – in a mirror-image pattern, which on the face of it was a clear display of red-blue equality. However, it was not always a successful one and, as time went on (each child spent between two and three hours on the game), nearly half of them tried out a version of a random arrangement as well. Several of those who made this move gave reasonably coherent explanations of why they did so, which showed that they had realised that a ball moving randomly in a random spatial arrangement would probably hit an equal number of red and blue balls.

These young children's success in spontaneously randomising the spatial arrangement of the various balls to achieve fairness (the researchers called this the 'unsteerable fairness' strategy) is truly impressive, and it seems to belie many of the negative conclusions drawn from the other pieces of research that we described earlier on in this section. If young children, or, at any rate, some young children, realise quite rightly that randomisation is an effective way of ensuring fairness in particular contexts, surely they understand randomness and the effects of randomisation a great deal better than all the other studies that we reviewed in this section suggest. What is the reason for the apparent difference in the implications of the Paparistodemou *et al.* study and the studies by Piaget and Inhelder, by Green and by Chiesi and Primi?

There are various possible answers. One quite plausible explanation for Paparistodemou *et al.*'s positive results is the time that the children spent on the problem and the opportunities that they were given to think about different strategies. The children in this study did not start by randomising the spatial arrangement: they eventually got there, for the most part, after having tried to solve the problem with symmetrical spatial arrangements. None of the other studies that we mentioned earlier seems to have given children much opportunity to change from one strategy to another.

Another possibility is that the study by Paparistodemou *et al.* was very explicitly about fairness, and the children who randomised may well have done so because of their previous experience of using randomisation to ensure fairness in games. The authors made the comment that the children's spoken comments were 'littered' with references to familiar games, which certainly suggests that those who decided in the end on randomisation did so because they already knew that this was a reasonable way, and sometimes the only way, of ensuring fairness in games. In our view, both explanations could be right, but we need a lot more research to find out whether they are or not.

This research would be extraordinarily interesting to do. One fascinating issue would be about how well children discriminate contexts in which randomness is the most effective way to achieve fairness and others in which it is not. We have said enough about the first kind of context: in the second kind of context, fairness can usually be achieved most effectively by some form of sharing. In fact, there is a great deal of research on children's understanding of

sharing as a way of distributing rewards or duties, and by and large this shows that even pre-school children understand why sharing works (Desforges and Desforges, 1980; Frydman and Bryant, 1988; Hay, *et al.*, 1991; Miller, 1984; Squire and Bryant, 2002). However, as far as we know, there is no research at all on how well children or adults distinguish the two kinds of context.

Randomness as uncertainty: the view of infants

One criterion of randomness, as we have remarked, is uncertainty. No one knows – everyone is uncertain of – exactly what will happen next in a random sequence. This uncertainty, on its own, is an insufficient criterion for randomness, since people may also be quite uncertain about what will happen next in a determined sequence, if they know absolutely nothing about what causes what in the sequence. Nevertheless, uncertainty about the next event is an absolutely necessary result of randomness, and it is therefore worth asking whether children recognise this link. Some interesting recent data on this issue has come from an unexpected source, from research on 10- and 11-month-old babies.

Xu and Denison (2009) set up a research study to answer the following questions. Do very young children discriminate between someone making a deliberate and informed choice and the same person making a choice blindly and at random? Do they understand that in one case the adult will choose the object that she wants, while, in the other, only the contents of the box will affect the probability of her pulling any particular object out?

To make this comparison, the researchers enacted a sequence of events, while the infant looked on, which started with them showing the infant some quite large boxes of red and white balls, in some of which there were more red than white balls and in others more white than red. Next, one of the researchers tried to make it clear that she preferred one colour over the other by picking balls of one colour only from a small container that contained both colours and showing her pleasure at having picked these particular balls. Then, the same adult researcher picked five balls from a large box whose contents the infant could not see at the time and had not seen before. She drew the balls from the box at different times in three different ways.

Two of these three conditions were *random-choice conditions*: they were designed to show the infant that the adult was drawing the balls at random without knowing what kind of ball she was choosing. In one *random-choice* condition, the adult simply looked away from the box and closed her eyes when she reached into it to withdraw each ball. In the other *random-choice* condition, she wore a blindfold while she pulled out the balls.

In contrast, the adult could see inside the box in the third condition, which we call the *informed-choice condition*. Each time that she picked a ball, she looked into the box through an opening in the top. From the point of view of the scenario that the experimenters devised, this meant that the adult could pick the colour that she preferred.

In all three conditions, the experimenters actually pre-arranged the colours of the balls that the adult drew from the box: it was always the case that the five balls were of one colour only, either all red or all white.

The point of the comparison between these three conditions was that someone who understands the consequences of randomness would expect differences between the third

condition, in which the adult preferred one colour and could make an informed choice which would be guided by this preference, and the other two conditions in which they could not see what they were doing when they reached into the box, and therefore could not deliberately choose the colour that they preferred. In the third condition, the adult should pick the colour that she prefers, while in the other two conditions the adult would be more likely to pick balls of the majority colour in the box itself. Thus, if she just picks red ones in the random-choice conditions, she is more likely to do so from a box that contains vastly more red than white ones rather than the other way round.

To see if the babies understood that the colour of the balls drawn by the adult might be different in the different conditions, the experimenters finally opened up the box, so that the infant could now see the contents of the box as well as the five balls that the adult had already drawn from it. The experimenters argued that in the two random-choice conditions, it would be surprising if the colour of the five balls drawn by the adult was not the same as the majority colour in the box, since it is less probable (though not impossible) that you would draw, for instance, five red balls from a box which contains mostly white balls than from a box which contains mostly red balls. On the other hand, in the (non-random) informed-choice condition, it would be surprising if the adult did not confine her choices to the colour that she preferred: thus if the adult preferred red, a choice by her of five white ones, when she could see inside the box, would be a surprise.

A large amount of research with infants has shown that they look at events that are in some way novel or surprising to them longer than they do at highly familiar and predictable events. So, Xu and Denison measured the amount of time that the infants looked at the scene once the contents of the box were revealed. They reported that in the random-choice conditions the infants looked longer (by around 2 seconds) when the majority colour in the box and the colour of the five balls drawn from it were different from each other than when they were the same. However, in the informed-choice condition the contents of the box made no difference to how long they kept looking at the event. In this condition, it was the adult's preference for a specific colour that made the difference: the infants looked longer (again by about 2 seconds) after the adult chose the colour that she didn't prefer than when she chose her preferred colour.

These results suggest that the babies were alert to the events that they witnessed and were able to make incisive inferences about the probability of what happened and about the personal intentions of the adult. The experimenters concluded that the babies understood (a) that the probability of randomly drawing a particular colour depended on the relative number of the two colours in the box, and (b) that a person making an informed choice would act on their own preference, and not on the relative number of red and white balls in the box.

Here, we will stick to their claim that these less-than-one-year-old infants understood that an informed choice would lead to one result and an uninformed choice to another. This is an important conclusion because, if it is right, it means that from an early age and even before they can speak human infants can discriminate some determined sequences from some random sequences for which the rules of probability applied. This view is, of course, strikingly different from the ideas of Piaget and Inhelder, and of some other researchers whom we have mentioned earlier in this section, about the age at which children begin to be able to distinguish random from determined events.

Is this new view justified? We will be discussing whether the pattern of the infants' responses in the two random-choice conditions really did indicate that they had made probabilistic choices in

a later section of this report. Here, we shall just express some caution about the informed-choice condition in the experiment that we have just described.

When the adult drew the five balls successively from the box, the infant could see the colour of each of the balls as soon as the adult took it out because she placed each of the five balls in a transparent container, which was in the infant's view. As soon as the balls were drawn out of the box and some time before the contents of the box were revealed, the infant who, according to the researchers, already knew which colour the adult preferred, could see whether or not the adult was choosing balls of the colour that she did prefer.

The researchers measured each infant's looking time from the moment when, later on, they revealed the contents of the box to the infant. This was the correct thing to do in the two random-choice conditions because, according to the researchers, the infant's looking time was determined by what the infant saw when he looked at the contents of the box. But, in the informed-choice condition, the relative number of red and white balls in the box, again according to the researchers, was irrelevant. The researchers claim that in this condition, the infants were surprised when the adult chose the five balls in a colour different from her preference and were not surprised when the colour that she chose was the same as the colour she preferred. The trouble, as we see it, is that the adult made her choices some time before the box's contents were revealed and therefore the act of opening the box and revealing its contents to the infant should not have had any effect on whether he or she was surprised or not. The fact that the babies, having seen a purportedly surprising event some time before the box was opened up, looked at the scene for a relatively long time some time after the box was opened up is, therefore, a puzzle. Given the importance of the researchers' conclusions, there is an urgent need for this puzzle to be cleared up.

Teaching randomness

We end this section, as we shall end subsequent sections, by considering what is known about how to teach children about randomness. This will have to be a very short section because there seems to be very little research on this obviously pertinent question. The nearest that we have come to an attempt to find some kind of an answer to it is the work on computer microworlds by David Pratt and his colleagues which we have already mentioned. Studies of microworlds are sometimes called 'teaching experiments' because the researchers concerned are interested not just in children's reactions at any one time but also in how their ideas change during the course of the study.

In fact, in these particular studies, much of what was written was about how individual children gained new insight into randomness during the course of the project. The authors claim that the main reason for these improvements was the interaction between the children's 'inner resources' and the technology of the microworld, which gave them an unusual degree of control over probabilistic situations. The inner resources were the basic intuitions that the children brought with them to the microworld, such as their idea of randomness as 'unsteerable', and the control was part of the microworld itself. In the Pratt and Noss study, the technology included what the researchers called the 'workings box', which allowed the children to generate random sequences and to regulate some aspects of these sequences, such as their size. Pratt and Noss argue that this aspect of the microworld played an essential part in the steps that the children took in understanding the idea of randomness during the study.

Thus, this study threw up some interesting suggestions about how to teach children about randomness. Yet, it was definitely not a full intervention study of the kind that we have just described – no pre- or post-tests and no control group. It would not be difficult to do such a study in order to test Pratt and Noss's hypothesis about the fruitfulness of interactions between children's intuition and their experiences of technological control, or to carry out other intervention studies to test other hypotheses about learning about randomness.

Summary

1. Although randomisation is a common and important part of everyday life, it is clear that even adults' grasp of the nature of randomness and its consequences is often tenuous. Many researchers have concluded on the basis of research on young children that they have even more difficulties in understanding randomness than adults do.
2. The research by Piaget and Inhelder (1975) on young children's predictions about successive randomisation led to the interesting idea that children first have to grasp the reversibility of determined spatial arrangements and then some time later go on to learn about the virtual irreversibility of progressively randomised sequences. However, the context that Piaget and Inhelder used for randomising in this study (a row of marbles rolling down a slope at the same time) was probably strange to the children, and the study needs to be done again with situations which children would be more familiar with like shuffling cards.

3. By the age of 10 years or so, children have a range of ideas about the nature of randomness, and even before that age they are able to some extent to make use of their association between fairness and randomness in a computer microworld. This association with fairness offers an excellent avenue to the study of children's understanding of randomness. Fairness engages their attention, and they seem to be able to adopt flexible and adaptive strategies to achieve fairness in different ways.

4. There is some evidence that children begin to understand the link between uncertainty, randomness and probability at a very early age indeed, but this is not conclusive. It would be easy to fill the gaps left by the interesting research that led to this important claim.

5. The overall picture that we have is that children are actively interested in randomness, particularly in the context of fairness, and that the ideas that they have about randomness should be encouraged through teaching.

3. Understanding and analysing the sample space

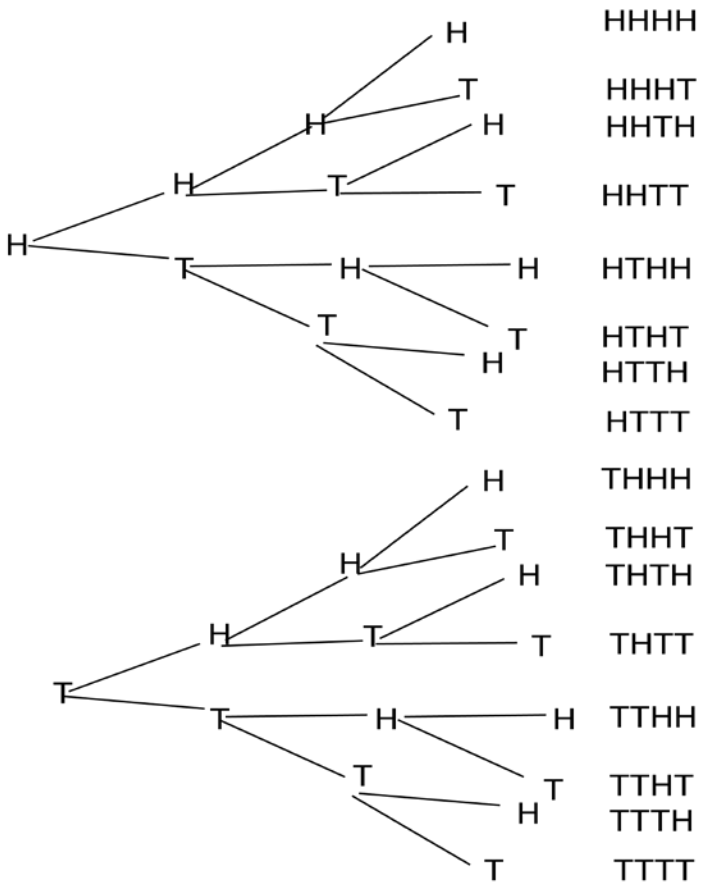
What is sample space?

Problems in probability are always about a set of possible, but uncertain, events that occur randomly. We cannot say what will happen next in these random sequences, but we can often try to work out the probability of particular events or particular types of event. To be able to do that, we need one crucial bit of knowledge - we have to know precisely what all the possible events are.

It is easy to see why this is so. Let us start with the simplest and most familiar examples – tossing coins and throwing dice. The probability that a tossed coin will land head upwards is 1 in 2, or 0.5, or 50% or ½, because there are only two possible outcomes of tossing the coin – a head or a tail. With the same kind of reasoning, one can easily work out that the chance of a die landing with 3 uppermost is 1 in 6 or 0.167, and it takes only a small step to realise that all the probabilities of all the possible events will sum to 1. The chance of throwing a head and of throwing a tail are both 0.5 which also add up to 1: the chance of throwing each of the 6 numbers on a die is 0.167 and the sum of the probabilities of the 6 numbers is again 1.
Of course, the range of possible events that one needs to know is rarely as simple as that. To continue with tossing a coin, what are the possible outcomes of tossing the coin twice, or three times or four times? It can surprise no one that the number of possible outcomes will increase with each extra toss of the coin. However, it is less obvious but just as important that the relationship between the number of tosses and the number of possible outcomes is not a simple linear one. It is in fact a multiplicative relation, and **Figure 1** (overleaf) shows why.

The left-hand side of the figure is a tree diagram that shows how the number of possible outcomes doubles on each throw. You can throw either a head or a tail with one coin, and each of these two possible outcomes has the same two possible sequels on the next throw, which makes four possible outcomes (2^2), and since each of these four possibilities can be followed by one of two possible sequels there are eight possible outcomes (2^3) on the third and 16 (2^4) on the fourth throw. Thus, the number of possible outcomes doubles each time that a coin is tossed and this is because the number of outcomes of each throw is two. A throw of a die has six possible outcomes and so the number of possible outcomes is multiplied by six with each extra throw: there are 36 (6^2) possible outcomes of throwing two dice and 216 (6^3) of throwing three.

Figure 1: *Tree diagram to represent the sample space for four tosses of a coin.*



We use a tree diagram in this figure, because we think that this is the clearest way to illustrate the sample space in random sequences of events. It is worth noting that this way of representing all the possible sequences in a sample space was used as a teaching instrument with some success in a study of children aged between 10 and 14 years (Fischbein, Pampu and Minzat, 1970).

The importance of the sample space

The term that is usually used for the number of possible outcomes in any probability problem is the ‘sample space’ or ‘problem space’ (Chernoff, 2009), and we have may have already said enough to show that knowledge of the sample space is an essential part of finding the correct solution to any probability problem. Every calculation of probability is based on the problem’s sample space. In fact, once one knows and understands the sample space, one is usually well on the way to solving the problem.

It is no surprise, therefore, that, some common confusions about probability and mistakes in probability problems can be traced back to many people having a hazy picture of the sample space of the problem that they are dealing with. For example, one common mistake is for people to expect random sequences to have no discernible pattern. Many adults judge BGGBGB as a more likely sequence than BBBGGG (Kahneman and Tversky, 1972) for the

birth order of six children in the same family. In fact, these two sequences are as likely as each other, as a glance at the sample space for this particular context would show. As with tossing coins, there are two possible outcomes, a boy or a girl, each time. Thus, the total number of equiprobable sequences for six births in a family is 2^6 or 64. BGGBGB is one of these 64 possible sequences and BBBGGG another, and so both have a probability of 1/64 or 0.016. The people who made this mistake had not been shown the sample space and may not have thought about it at the time. It seems quite likely to us that they would change their minds quite quickly if they were told about the 64 possibilities and grasped the fact that these are all equiprobable.

Another common mistake, which we have described already, might also become much less common if people thought clearly about the sample space when trying to solve probability problems. This is the negative recency bias (Chiesi and Primi, 2009), which leads people to predict a change after a run of one particular event. They judge a head as more likely than a tail on the next toss after a run of tails even though the two events are equiprobable. If they looked at the sample space for tossing a coin, for example, five times they would see that there are 32 possible equally probable sequences, of which one is HHHHH and another HHHHT. Since each of these sequences is equiprobable – these two sequence both have a probability of 1/32 – one is just as likely as the other. So, the sample space is a vivid and, we imagine, effective way of showing people that recency (what happened on the last four throws) is not a good basis for a judgement about probability.

We will be making the point that the crucial importance of the sample space tends to be neglected both in research on children's understanding of probability and in teaching children about probability but before that we will present one other example of a fairly sophisticated kind of judgement about probability which people often get radically wrong even though they would probably see through their own confusion by thinking clearly about the sample space.

The probability of a particular event happening to you in the future depends on the number of occasions on which it could happen. The probability that we, the authors, who live in middle England, will meet a man wearing a kilt at our local shopping centre next week is fairly low, but it will be greater if we go there three times than only once during the week. The probability that you will throw a head is higher if you toss a coin three times than if you toss it twice or only once, as **Figure 1** clearly shows. This pervasive and completely reliable relationship between the two variables that we are talking about – (1) the probability of a particular event and (2) the number of occasions on which it could take place – could easily lead people to think that the relationship is a linear one. Throw the coin three times and you will be three times as likely to throw a head as when you throw it only once.

This idea is wrong. There isn't a linear relationship here, as a glance at **Figure 1** shows: the probability of throwing a head when you toss the coin once is 0.5: if a linear relationship did hold, that probability would double (p = 1) with two tosses and would triple with three tosses (p = 1.5). Since a probability of 1 means absolute certainty, that would mean that you would get a head at least once by tossing the coin twice, which is manifestly not the case, and the figure of 1.5 for three tosses makes no sense at all. In fact, as **Figure 1** shows, the probability of one or more heads in two, three and four throws is well below certainty. There are four possible outcomes in the sample space for two-coin tosses, and eight for three-coin tosses and 16 for four. In each sample space, only one of the outcomes includes no heads at all (TT for two, TTT for three and TTTT for four tosses). Thus, the sample space tells you that the chance of getting

at least one head in two throws is 3 in 4 or 0.75, in three throws 7 in 8 or 0.875, and in four throws 15 in 16 (0.94). Thus, the relation between the number of throws and the probability of this particular event is not a simple linear one: you are not four times more likely to get at least one head when you toss the coin four times as when you toss it once.

However, when a group of Belgian psychologists (van Dooren,*et al.,* 2003), asked 16- and 17-year-old students to judge the truth of a set of statements about throwing dice on a certain number of occasions, most of these students seemed to assume a linear relation between the number of times the dice were thrown and the probability of particular outcome, even though the older (but not the younger) students in this group had already been formally taught about probability at the time of the study. The statements that the students were asked to judge were a great deal more complex than the coin tossing example that we have just outlined, but the principle was the same. One, for instance, was: *I roll a die 12 times. My chance of getting at least two 6s in these 12 throws is three times as great as my chance of getting at least two 6s if I roll the die four times,* and then the participant had to tick the box beside one of two answers *This is true* or *This is not true.* The statement is a claim for a linear relationship between the chance of a particular outcome and the number of times that the dice are thrown, and it is not true: the probability of rolling at least two 6s with 12 throws is more than three times the probability of the same outcome with only four throws. Yet, the great majority of the students who took part in this study and even of those who had been taught about probability, opted for the linear relation most of the time in this and some other similar problems.

This is an understandable mistake. The solution to many of the mathematical and scientific problems that students are asked to solve at school lies in a linear relationship between two variables, and it is not at all surprising that these students tried the same approach in probability problems as well. Yet, it seems to us that they might have been far less likely to do so if they had been encouraged to think about the sample space. This, too, would show quite clearly that the relationship between the number of possible sequences with two or more sixes in them and the number of times that the die is rolled is complex and certainly not linear.

We are speculating here about a possibility that needs investigation. In fact, our speculation suggest two sets of studies.

1. Students may make the linear error because they have no clear idea of the sample space, but we cannot be certain that they would know how to interpret and use the sample space. A study in which the researchers provide the students with a clear account of the sample space for some probability problems but not for others should sort this out. If our speculation is correct, those given a clear and comprehensible sample space for problems like the ones that van Dooren *et al.* gave to students, would no longer make the linear error. Exactly the same comparison could be made with the negative recency effect, which we described in the previous section. The sample space for three-coin tosses consists of 8 (2^3) equiprobable sequences, and a list of these would easily demonstrate to the student that the sequence HHH is just as probable as the sequence HHT and therefore that one is just as likely to throw another head after throwing two heads already as to throw a tail the third time round.

2. The next question is how to teach children to work out the sample space for themselves. As far as we know, there is no research on this issue, apart from some studies that we shall review later on children's ability to categorise material systematically. Research on

teaching children about creating sample spaces should concentrate on how best to present the sample space and on how to convince children of the need to list all the possibilities exhaustively. One way of approaching the first issue is to test Fischbein's suggestion that tree diagrams are an effective way of teaching children about how to list all the possibilities, by comparing the use of tree diagrams, like the one on the left of **Figure 1** with the use of exhaustive lists such as the one on the right of the same figure. The complete lack of systematic intervention studies on methods of teaching students' how to form sample spaces is, in our view, the most glaring gap in research on children's understanding of probability.

Aggregating the sample space: working at two levels simultaneously

In some problems the relation between the sample space and the probability of a particular outcome is a simple and quite a straightforward one. Since there are eight possible outcomes when you throw a coin three times and they are all equally probable, the chance of one outcome happening is 1 in 8 (0.125). In many problems, however, the relation is more complex than that, because the people solving the problem must first group or categorise the sample space in some way before they can reach any conclusions about the probability of particular events. Usually, this is because the outcomes that are relevant or important are formed from combinations or categories of the possible events in the sample space. Usually also, as we shall see, the new combinations and categories are not equiprobable even when the basic elements in the sample space are.

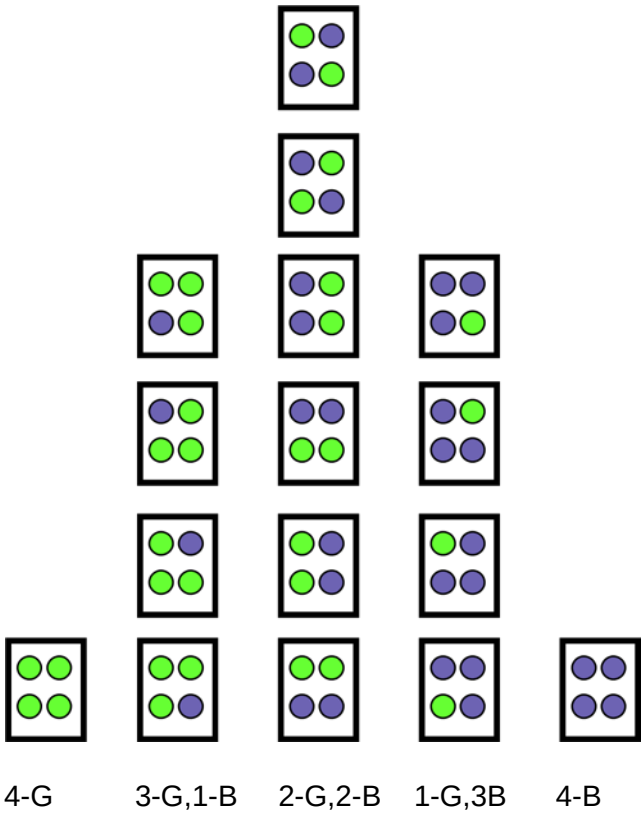
In the introductory section, we mentioned university students' difficulties with the DM problem devised by Keren (1984). Recall that the students were told about two boys, D and M, who together drew three cards from a pack having an equal number of red and black cards. If the card was red, one of the boys won a coin, and if it was black, the other boy won. The students were asked about the probability of two kinds of event: one was that one boy would win all three times, the other that one boy would win twice and the other once. The correct answer is that second of these two categories is much more common than the first: this is because there are only two possible ways in which one boy could win all (DDD or MMM) but six possible ways of one boy winning twice and the other once (DDM, DMD, DMM, MMD, MDM or MDD).

This should be an easy problem to solve for anyone who has looked at the sample space, and so it is very likely that the reason for 48% students producing the wrong answer (most of these said that there was an equal likelihood of the two kinds of event) was that they did not think clearly enough about all the eight possible outcomes. One possible reason for this might be a confusion caused by the need to think about two levels of analysis at the same time. One level is the two categories that the participants are being asked about, and the second level is the eight separate outcomes from which the two categories must be formed. Even though they combine the individual outcomes to form the two categories, they still have to keep them separate in order to realise that only two of the outcomes belong to one category (one boy wins every time) and the remaining six to the other (one boy wins twice and the other once). This is the information that they need to be able to judge the second category as more probable than the first. However, the eight individual possible outcomes that they have to consider are equiprobable, and this is what might have led the participants to say that the categories were equiprobable too.

We cannot be sure that this is the underlying problem and the answer must lie in further research on the effects of providing participants with information about the sample space which either clearly distinguishes the two levels of analysis or not. This approach would be also be valuable to help decide the implications of some further research, this time with children that also involves aggregation of categories which are unequal in their probability but which are formed from basic elements that are equiprobable.

This is a set of studies by Abrahamson (2006, 2009). His method was to ask children what would happen when they used a scoop to collect four balls at a time from a container that held a large number of green and blue balls. The scoop was a flat square with a handle and with four holes in it each of which was the right size to hold one of the balls. So, when the children plunged the scoop into the container and then retrieved it, the scoop would usually hold four balls, which were either blue or green ones. This meant that there were 16 possible equiprobable outcomes for any scoop: BBBB, BBBG, BBGB, BGBB, GBBB, BBGG, GGBB, BGBG, GBGB, BGGB, GBBG, GGGB, GGBG, GBGG, BGGG and GGGG. This is all one would need to know about the sample space to answer a question about the probability of one particular outcome, such as BGBG. The answer of course is 1/16. But Abrahamson's questions were not just about individual outcomes such as BGBG, but also about categories of outcomes. He wanted to know whether children could work out the probability of the scoop holding, for example, just green balls or just blue balls, or an equal number of blues and greens. The correct answer is that the probabilities of these three categories are very different, as **Figure 2** shows. There are six different ways in which the scoop could hold two blue and two green balls, but only one way in which all the balls would be green or all would be blue. Thus one is six times more likely to scoop an equal mix of greens and blues than to scoop just one colour.

Figure 2: *The 16 possible outcomes in Abrahamson's 4-scoop problem.*



As in Keren's study with adult students, so in Abrahamson's research with school children, the participants had to deal with the same sample space at two levels simultaneously. They had to be able to work out each of the five categories, but in order to compare the likelihood of these different categories they also had to be aware of the 16 individual possible outcomes since the categories were formed from different numbers of these outcomes. To put it in Abrahamson's terms they had to deal with and distinguish 'aggregate' and 'elemental' events. An elemental event is, for example, a scoop with one blue ball in the top right hand corner and three green balls in the other available holes, whereas an aggregate event is a scoop with three green and one blue ball in any arrangement. The crucial point of Abrahamson's task is that the 16 elemental events are equiprobable but the aggregate events are not, and his main question was about how well children could understand this difference.

Abrahamson's answer to this question was based on qualitative reports of the often-changing answers of children around the age of 12 years. He reports that it is hard for them to coordinate these two levels of data with each other. Sometimes, they judge that two aggregate events are equiprobable when in fact they are not, and their justifications for this response make it clear that they think of the aggregate event, such as three greens and one blue as an elemental one. This is not a surprising error since two of the five aggregate events (four blues, four greens) are elemental events too which must make it hard for the children to distinguish the two levels. At other times, the student will argue that some elemental events are more likely than others. When a child does make this mistake he or she usually claims that an elemental event, which is part of a relatively frequent aggregated category, like two green and two blue balls, is more likely to happen than another elemental event, like four green balls, which belongs to a much less common category.

The mistakes that we have described so far seem to be due mainly to children paying little or no attention to the sample space. The students' answers that Abrahamson reports suggest a different kind of difficulty. Abrahamson, with the help of various graphical representations, encouraged the children to think about all the possible events and their relative frequencies. Thus they had at their fingertips clear information about the sample space. Yet, even so, they sometimes confused the two levels of analysis.

Compounds and aggregations

To solve some probability problems, the aggregation that is needed is a matter of combining two elements. In some experiments the participant are asked about the relative probability of these combinations or compounds. This was what happened in a series of experiments by Lecoutre (1992; see also Lecoutre and Durand, 1988). In these studies Lecoutre posed the following problem, in various different versions, to a large number of adults and also to groups of secondary school children: *There are three poker chips in a box, two red and one white, and two are drawn (from the box). Is it more likely (a) that two red chips are drawn or (b) that one red and one white chip is drawn, or (c) are these two events equally likely?* The sample space for this problem contains three possible outcomes:

1. that two red chips are drawn
2. that one of the red chips and a white chip are drawn and
3. that the other red chip and a white chip are drawn.

Thus, there is only one way to draw two red chips, but two different ways to draw a red chip and white one. So, the second and third of the three possible events form the aggregate event of one red and one white chip, and this compound event is twice as likely to happen as the event of drawing the two red chips.

Some of the people to whom Lecoutre and her colleagues gave this problem had had several years of teaching about probability, others not. Some were science students, others arts students. Some had had a great deal of more experience than the others with games of chance. The main result of this research, however, was that around 50% of the participants gave the wrong answer, and in nearly every case their mistake was to judge the two events (two red chips versus one red and one white) as equiprobable.

The kind of formal education that the participants had had seemed to make very little difference to this pattern of results. Arts and science students did equally well, and those who had taken courses on probability did no better than the students who had not. The people who were highly experienced in games of chance did make the mistake noticeably less than those with less experience of this sort, but even so very nearly 50% of them gave the wrong answer. This was about the simplest possible aggregation problem (only three possible elemental events and an aggregated event formed from two of these events) and yet the majority of the participants did not manage to solve it, even though many of them were very well educated.

Lecoutre then tried another version of this task, this time with children. In her new task the participants were given three geometrical shapes instead of poker chips. Two of these were identical triangles and the third was a square. They were shown, before the task began, how they could put the square and one of the triangles together to make the shape of a house if these were the two shapes that they drew out, whereas drawing out two triangles would allow them to construct another geometric shape (a rhombus) and not a meaningful figure. They were then asked the same kinds of question as in the poker chip task: *If you draw out two of the three shapes at random, are you more likely to draw out a rhombus, or a house, or are these two events equally likely?* The same children were also given a second task which was a virtual repeat of the poker chip task except that the objects were three sweets, two of which were orange- and the third lemon-flavoured and they had to judge the relative likelihood of picking one orange and one lemon, or two orange sweets.

Lecoutre found a very different pattern of results in the two tasks. The children did much better with the house-rhombus problem than with the orange and lemon sweets. In the first task three-quarters of the group answered correctly that the house was a more probable event than the rhombus: only 23% made the mistake of judging the two events to be equiprobable. This of course is a much more successful set of answers than those of the adults in the poker chip task, even though the structure of the two tasks was the same However, in the second task with the orange and lemon sweets roughly the same numbers of children made the wrong, equiprobable choice as made the correct choice, which is only a slightly higher rate of success than in Lecoutre's previous studies.

What made the house-rhombus problem a relatively easy one? Lecoutre's own explanation is that this problem distracted the children from the random nature of the choices by making them think more about the construction of the shapes and about the material that they had to hand to make these constructions. This seems an unconvincing explanation to us since the questions that the experimenters put to the children in this task could only make sense to them if they

understood that picking two shapes could lead to quite different outcomes. To get the right answer the children had to understand that chance determined which shapes they would draw out. Their relative success in this task must have been due to them being able to see that there were two ways of making a house, because there were two different roofs, whereas there was only one way of making a rhombus. The use of a concrete, identifiable and familiar object like the roof of a house made it easier and more natural for the children to analyse the sample space logically and correctly. The children's success in this task shows that there are contexts in which most of them see the need for analysing the sample space correctly, and carry out the analysis correctly. The educational implications of this conclusion are great, and should be taken up.

Questions about compounds can be more complex than the one that Lecoutre posed. There is for example the quite commonly given *two-dice problem* in which the child is asked about the different possible totals that one could get when throwing two dice at the same time and then adding the two numbers you have just thrown. **Table 1** shows how, when two dice are thrown, one of 36 possible compound outcomes will be the result. In the different compounds, the two dice add up to eleven possible totals that range from 2 to 12. It is easy to see from the table that the number of compounds that sum to these different totals varies a great deal between totals. There is only one way of throwing a total of 2 or a total of 12, and there are two ways of throwing the totals 3 and 11, three ways for the totals 4 and 10, four ways for the totals 5 and 9, five ways for the totals 6 and 8, and six ways for the total of 7. To take one example, you are six times more likely to throw a total of 7 than a total of 12. Thus, this is another context in which the elements, which in this case are compounds, are equiprobable, but can be aggregated by the sums of two numbers in the compounds and these aggregates are not equiprobable.

Table 1: *The 36 possible totals made by summing the outcomes of throwing two dice (A and B) at a time.*

		The number on which A lands						
		1	**2**	**3**	**4**	**5**	**6**	
The number on which B lands	**1**	2	3	4	5	6	7	**The sums of the two numbers**
	2	3	4	5	6	7	8	
	3	4	5	6	7	8	9	
	4	5	6	7	8	9	10	
	5	6	7	8	9	10	11	
	6	7	8	9	10	11	12	

These large and systematic differences in the probability of the sums of two dice are, it seems, not immediately obvious to schoolchildren. Fischbein and Gazit (1984) included some questions about throwing two dice in a questionnaire that they gave to 10- 11- and 12-year-old students. Among other questions, they asked the students what the probabilities were that that the sum of two dice would be (a) 6, (b) 13 (an impossible event) and (c) bigger than 9. The children did reasonably well in spotting that 13 was not a possible option. Thirty-eight percent of the 10-

year-olds, 81.0% of the 11-year-olds and 78.8% of the 12-year-olds got this question right, which means that they understood the question and were aware of the limits of the sample space. However, the number of correct answers to the other two questions, particularly by the youngest group, the 10-year-olds, was very low indeed. None of the 10-year-olds, but 51.0% of the 12-year-olds, came up with the correct answer to the question about the probability of the total 6 (question (a)), and none of the 10-year-olds, and only 29.8% of the 12-year-olds, managed to work out the probability of a total greater than 9 (question (c)). Why was it so difficult to find an answer to these apparently straightforward two questions?

Perhaps the best way to solve this puzzle is to reflect on what the correct answers are, and how to find them. We can start with the total number of compound events, which you can see from **Table 1** to be 36. Since there are five ways of throwing the total 6 (1,5; 5,1; 2,4; 4,2; 3,3), the answer to question (a) is that the probability of this particular total is 5/36 or 0.139. A glance at the triangle of figures at the bottom right-hand corner of the collection of totals in **Table 1** shows that there are three different ways of throwing the total 10, two ways of throwing the total 11, and one way of throwing the total 12, which means that the answer to question (c) is 3+2+1/36 or 0.167. So, the problems are quite easily solved, provided that one has a version of **Table 1** to hand, and knows how to use it, because the table establishes the two crucial pieces of information: first the 36 possible compounds that can be thrown and second the number of these compounds that add up to the various totals in question. Fischbein and Gazit (1984) claim that many of the students' mistakes were due to them thinking that the number of possible compounds in the sample space is 12 not 36, and these researchers go on to suggest that this shows that the pupils were not using multiplicative reasoning because they added 6 to 6 when they should have multiplied 6 by 6 to reach the correct total of possible compounds.

The need to multiply is certainly one possible obstacle, but there are others. Another is what is known as the 'equiprobability bias': this is the assumption, which we encountered already in the results of Lecoutre's research, that, in a random arrangement, all possible events, including aggregated ones, have the same probability as each other. A third is that sample spaces, being multiplicative, quickly proliferate and, thus, are often hard to comprehend and to calculate as well.

One interesting solution to this last problem to work with computer microworlds, in which it is possible and quite easy to simulate the results, for example, of tossing two coins or two dice 1000 times. A study by Pratt (2000), which is about 10-year-old children's solutions to the two-dice problem in the context of a computer microworld, is a clear illustration of the advantages of this kind of simulation (see also Konold, Harradine and Kazak 2007).

Pratt's study confirmed that 10-year-old children sometimes do assume equiprobability in their first attempts to solve the two-dice problem. They argue that all the 11 possible totals are equally probable. Pratt worked with children in a computer microworld in which they could study for themselves the results of a very large number of throws, for example 1000 throws of the two dice. He noted that the children tended to begin working in this microworld with the assumption of equiprobability, which they adopted because of the strong association that they had between equiprobability and randomness. Later, however, partly as a result of experiences that they had of manipulating aspects of the microworld, and also of seeing the results of spectacularly large numbers of throws of the two dice, they exchanged this assumption for a more appropriate view of the probabilities of 11 different aggregated totals.

There may be another possible obstacle here, which is that the compounding in the two-dice problem requires a detailed knowledge of the additive composition of number, which is the knowledge that numbers are composed of other numbers: for example 7 is composed of 1 and 6, 6 and 1, 2 and 5 and so on. Even though quite a lot of evidence (Nunes and Bryant, 1996) suggests that children do have some understanding of additive composition quite some time before the age of 10 years, the detailed knowledge of composition that the two-dice problem requires may be too difficult for them at this age.
At the moment there is little systematic evidence to tell us what the reasons for children's difficulties with this compounding problem are. In our view, there is a need to find out whether the cognitive obstacles that cause whatever difficulties children have with any probability problem are general ones or are specific to probability.

The role of combinatorial reasoning

Piaget and Inhelder (1975) claimed that combinatorial reasoning is a fundamental part of children's learning about probability. They argued that children begin to understand the nature of randomness through being able to work out all the possible combinations in random situations. 'The child constructs his notion of probability by his ability to subordinate the disjunctions effected within mixed sets to all the possible combinations, using a multiplicative and not simply an additive mode.' (p. 161). In other words, children make sense of randomness by working out all the possible combinations in the sample space. Piaget and Inhelder's hypothesis was that children only begin to analyse combinations systematically at around the age of 11 or 12 years, and so they concluded that the need for combinatorial reasoning is a major obstacle to children's understanding of probability.

This claim is plausible enough and it certainly justifies the large number of studies on combinatorial reasoning that Piaget and Inhelder report in their book on children's ideas about chance. One of these studies was directly about combinations and probability. The ages of the children in this study ranged from 5 to 12 years. The experimenters first showed each child a set of counters on the table in front of them. These came in four different colours, and the number of counters in each colour varied (e.g. 15 yellow, 10 red, 7 green and 3 blue). The experimenters then put an exactly identical set of counters into a sack and shook this sack thoroughly. Then, several times over, they asked the child to draw out a pair of these counters, but also asked her each time to make a prediction first of the colours of each counter in the pair that she was about to retrieve. The counters that the child drew each time were not returned to the sack, and thus the probabilities of the possible pairs varied constantly throughout the session. So, in order to calculate or estimate the probability of each possible combination of colours in this rather complex task, the children had to take into account not just the original number of counters in each colour, but also the effect of the constantly changing sample space.

Piaget and Inhelder (1975) reported that the younger children (mostly 5-, 6- and 7-year-olds) showed no sign of any systematic analysis of the probability of drawing the different colours and often did not even take into account the original numbers of each colour. Slightly older children base some of their judgements on the initial numbers but did not monitor the changing sample space. The oldest children, around 10, 11 and 12 years in age, did make use of both kinds of information, and their predictions were realistic, as a result, and often successful. Piaget and Inhelder's conclusion from this developmental pattern was that children in this age range learn not just to estimate probabilities of certain events but also how to reason about combinations, and they argue that the children's growing ability to understand and to imagine combinations is

actually the basis for their eventual understanding of chance: ‘The essential conclusion to be drawn from the preceding observations is that the notions of chance and probability are by nature essentially combinatoric’ (p 128).

The idea of the central importance of combinatorial reasoning in learning about probability is plausible and exciting, but Piaget and Inhelder’s enthusiastic conclusion about the understanding shown by the older children seems to us to go too far. The predictions that these children made and the justifications that they gave for their predictions do show that they did take into account the relative number of each colour at the start and also the changes in the original sample space. However, the two examples given in the book of individual children making rational and systematic predictions do not establish that, as Piaget and Inhelder seem to be claiming, these children were able to work out all the possible pairs that they might draw in their next choice. One child, who consistently predicted a white and a red pair when white and red were the two most numerous categories, may simply have done so on the basis of the relative frequency of the individual colours rather than by working out all the possible paired combinations between them and estimating the relative frequency of all these different pairs. The other child consistently predicted a pair of green counters, when green started as and remained the most numerous colour, and again this prediction could have been based on the relative frequency of the individual colours and not on the relative frequency of possible pairs of colours. There is no evidence that this particular child even considered the possibility of drawing a combination of two different colours.

It is difficult to understand Piaget and Inhelder’s clear optimism about these children’s ability to analyse the sample space and to monitor its changes. Their reasoning seems circular. They argue that combinatorial reasoning is essential to understanding probability, and then conclude that a modest improvement with age in the success that children have in this probability problem must be due to an increase over time in children’s ability to reason about combinations.

Cartesian product problems

We have concentrated so far on children forming pairs either as compounds or just as combinations, but of course there are other kinds of combinatorial tasks, such as permutation tasks and arrangement tasks. Piaget and Inhelder (1975) did give such tasks to children, but the experiments that they reported were not directly connected to the topic of chance. The authors included these further experiments in their book on chance because of their assumption of the great importance of combinatorial reasoning in solving probability problems, but as we have already remarked this is an assumption that still needs to be tested.

One kind of combinatorial problem that has attracted a great deal of interest more recently is the Cartesian product problem. In this, the participant has to work out how many combinations are possible between two different kinds of material. In a typical example, the question is how many different kinds of sandwich can be formed from three kinds of bread (white, granary, wholemeal) and four possible fillings (tuna, cheese, avocado, chicken). It is well documented that this multiplicative problem is a difficult one for children younger than 10 years, who often add the relevant numbers (e.g. 3 + 4) instead of multiplying them (Brown, 1981; Nesher, 1988). The person who makes this kind of additive error is clearly far from creating an exhaustive list of all the possible combinations.

However, two different studies suggest that quite young children often can solve Cartesian product problems if they are allowed to model the situation with concrete material. Lynn English (1991), an Australian psychologist, gave children, who were aged from 4 to 9 years, a product problem about two kinds of clothing – tops and skirts or tops and trousers. The children had to work out what all the possible combinations of tops and skirts of trousers were. English also gave them teddy bears and encouraged the children to dress them before they decided on the possible combinations. English reported that the children's approach to this problem became more systematic and exhaustive with age. The oldest children (9-year-olds) formed all the appropriate combinations on more than half the times. In the study the children were given several different problems, and English noted that a marked improvement, mainly in the older children, in how systematic and successful their solutions were during the course of the study. She attributed the children's surprisingly high rate of success largely to the opportunity that they had to use concrete material to model each problem. This is a plausible, but untested claim. We need a study that compares children's solutions to Cartesian product problems with and without the help of concrete material.

In a closely similar study (Bryant, Morgado and Nunes, 1992; also described in Nunes and Bryant, 1996, p163–165), we also gave 8- and 9-year-old children concrete material to model a Cartesian product problem about clothing, again (co-incidentally) about combinations of t-shirts and shorts. One group of children was given all the material that they needed to model all the possible combinations. A second group was given an incomplete set of material: it included examples of the t-shirts and the shorts, but not all the different elements in the problem, The aim of this second condition was to find out if the children could work out what all the combinations were without being able to get to the solution by counting actual combinations that they had formed with the material provided.

We found that the 8-year-old children solved the problem successfully about a third of the time when they were given a complete set of materials but hardly at all with the incomplete set. The 9-year-olds did better. They calculated the number of possible combinations correctly just over half of the time with the complete set and roughly a third of the time with the incomplete set. So, it is generally hard for children under the age 10 years to work out a complete list of all the possible combinations in a Cartesian product problem, but it does helps children to have concrete material to hand to help them to think about these combinations. Again, we need more research to see what implications these two studies of children's solutions to Cartesian product problems have about children's understanding of probability and, in particular, about their ability to create an exhaustive and appropriate sample space for themselves.

Imagining possibilities

The analysis of sample space starts with an exhaustive search for all the possibilities that fall within that space. This makes two obvious intellectual demands. One is to eliminate from the space any element that is impossible. The second is to draw up the list of all the events that are possible.

There is more research on the first than on the second of these two demands. On the whole, researchers have reported that children quite easily eliminate impossible events from the sample space. Piaget and Inhelder report, in their chapter on the comparisons of two probabilities, that children solve problems in which one of the sample spaces has none of the quantity that they are being asked about (no blue balls when the question is about the relative

probability of drawing a red ball) relatively easily. As we mentioned earlier (page 42), Fischbein and Gazit also report that a question about an impossible total in the two-dice task was much easier for 10-year-old children to answer than other questions about events that were possible. In research outside the topic of probability, there is a large amount of evidence that even very young children are quite good at discriminating impossible events from ordinary events (e.g. Chandler and Lalonde, 1994; Subbotsky, 2004; Woolley and Cox, 2007; Johnson and Harris, 1994).

However, two fairly recent studies by Shtulman and Carey (2007) and Shtulman, (2009) show that the discrimination between possible and impossible events is not always that easy for young children. It is difficult for them, these researchers claim, when they have to compare and discriminate possible but highly improbable events and events that are completely impossible. Their studies showed that children between 4 and 7 years are quite good at judging impossible events, like catching a shadow, as impossible, but tend to include improbable events, such as catching a fly with a pair of chopsticks, in the impossible category as well. Children of 8 to 9-years, on the whole, did better on this task, even though some of them still did confuse improbable with impossible events. The researchers make no connection between these results and children's understanding of chance, but there is plainly a connection to be made since many of the serious risks in children's lives involve quite possible, but statistically improbable events, such as catching AIDS or death from a drug overdose.

There is, as far as we know no direct research at all on the second question. We simply do not know how children set about imagining all the possibilities in a particular sample space or what difficulties they have in doing so. There is some research on how well children reason about future events. This is currently the subject of some lively and interesting research (Atance and Meltzoff, 2005, 2006; Suddendorf and Corballis, 1997, 2007; Suddendorf and Busby, 2005; Russell, Alexis and Clayton, 2010), but we will not describe it in detail because it deals entirely with deterministic contexts. In these studies, children are asked to work out, for example, what they will need in order to play a particular game or to make sure that they are not bored in a particular situation. The results of these studies suggest that this is something that children of about 5 years generally can already do quite well.

However, we do not yet know much about how children set about forming a complete list of possible future events whose probabilities vary from high to low, and we know nothing about how to help them do so. This seems to us to be a gap in research on children's understanding of probability.

Summary and comments on teaching children about sample space

There can be no doubt of the crucial part that the sample space plays in our dealings with probability. Without a thorough and exhaustive grasp of all the possibilities involved in any probability problem, one has no chance at all of solving the problem except by applying some ill-understood procedure. Yet, we have very little direct evidence about the impact of children's knowledge of the sample space on their understanding of probability.

Two kinds of research study are needed very badly. One, which we have mentioned several times already, would be to see how much it helps children who are trying to solve probability problems to provide them with the relevant sample space. If this does improve their success in solving probability problems, it would be reasonable to conclude that children can see the point

of, and can take advantage of, knowing about the sample space, but for one reason or another are not managing to create it for themselves. It would then be important to find out what these reasons are. However, one possible result of this line of research might be that just presenting children, particularly quite young children, with the sample space may be no help at all. They may also need to be shown how to interpret the sample space as well. Whether or not they will need this additional help is an interesting and important question, which is quite easy to answer.

It is only a short step from this first line of research to the second set of studies that we think should be done, which would be about effective ways of teaching children how to form the sample space for themselves. In fact, the intervention methods that one would look at in this research would partly depend a great deal on what is found in research on the effects of providing children with the sample space on their solutions to the probability problem. The work that we have reviewed on the help that children seem to get from modelling possibilities with concrete material suggests that such material should be at the centre of teaching them about how to create a sample space.

One advantage of doing intervention experiments is that they would also, at last, test the as yet untested assumption shared by Piaget and Inhelder and by Fischbein and by many other researchers, including ourselves, that many of the difficulties that children have in probability problems are directly due to a failure on their part to create an adequate sample space. If this assumption is right, then successfully teaching children how to form the appropriate sample space should also improve their success rate in probability problems.

Several research teams have measured the effects of interventions designed to help children to form a comprehensive sample space. The interventions by Jones, *et al.,* (1997, 1999) with 8-year-olds and by Polaki (2002) with 9- and 10-year-olds are the best examples. The methods used in these studies were designed to encourage the children to form the sample space themselves and both studies report some success. However, neither research team included a control group or involved pre- and post-tests in their research, and so we cannot draw conclusions about the intervention methods with any confidence.

We have found one study by Barrattt (1975) of the effect of intervention on combinatorial reasoning which did include a pre-test and both an immediate post-test given soon after the intervention was over and a delayed post-test two months later. The researchers compared the answers of 12- to 14-year-old children in an experimental and in a control group in these tests. This well-designed study was on the question of how to teach children about combinations in a general sense: Barratt does not seem to have involved probability explicitly in the problems that he gave to the participants, but the results of the study are clearly relevant to the understanding of uncertainty. He gave all the participants two teaching sessions. The children in the experimental group were encouraged to use concrete material (e.g. dolls, counters) to solve a series of combinatorial problems, and when they had worked through each problem they were asked to check their answer against a correct and systematic solution that they were given. The control group children were given other mathematical tasks. Barratt found an impressive improvement from pre- to post-test in the oldest children in the experimental group, and less of a change in the control group children of that age. There was little difference between the two groups at any stage of the experiment in the younger children. The result suggests that it is possible to teach older children quite effectively about how to form combinations. It is possible that the younger children too might have benefitted from a longer training programme.

We are surprised to find no intervention studies designed to help children with another aspect of forming the sample space. Creating a sample space is usually a two-step process. First one lists the basic elements. Second, one aggregates these elements to form new variables, but even when aggregating elements one has to keep them separate from each other as well, in order to be able to work out the probability of the aggregate events. There is ample evidence to show that this is often a problem for young children for two different reasons. One is that the aggregation itself can be quite hard for children since it often depends on combinatorial reasoning that children find hard anyway. The other is that the need to combine events into aggregates and yet keep them separate is difficult for young children, particularly when the elements are equiprobable and the aggregates are not. Again, there is a sore need for research on teaching children, this time teaching them how to act on both cognitive levels (elements and aggregates) at once: we suspect that the use of concrete material to form models of the problem would be an effective help here too.

Finally, we wish to make the point that there are many reasons why children and adults often do not manage to form an appropriate sample space for themselves. Piaget and Inhelder concentrated on the need for combinatorial reasoning. This certainly must be an important part of children's difficulties with sample space, but there are probably other difficulties as well and these need to be researched. The cognitive demands of forming a sample space vary from problem to problem. For example, success in the two-dice problem depends on the participants' understanding of the additive composition of number at least as much on their ability to form compounds. At a more basic level, one has to be able to work out what is possible, even though improbable, and what is entirely impossible. As we have seen, this is not something that we can take for granted in young children.

So, it seems to us that much depends on research in the future on children's learning about sample space. It is quite easy to see what research should be done, and quite easy to do the research itself. This research would be a major step towards understanding how children learn and can be helped to learn about probability.

4. Quantifying probability

Probability and proportions: probability as an intensive quantity

Probability is a quantity, and learning how to calculate the probabilities of various events is an essential part of understanding chance. The information that we use to calculate probabilities always comes from the sample space, and the calculation that we perform on this information is almost always a proportional one. That is because probability is a proportional quantity or, to use the correct technical term, an intensive quantity.

All quantities can be categorised either as extensive or as intensive. Many of the quantities that children deal with in the context of mathematics are extensive quantities like mass, height, distance and the number of objects in a set. These extensive quantities obey simple additive laws. If I add a kilo of apples to the shopping that is already in my shopping bag, I increase the mass of its contents. The longer the tap is on, the more water collects in the basin.

This is not the case with intensive quantities. If the temperature of the litre of water in the basin is 20°C and I add another litre of exactly the same temperature to it, the (extensive) amount of water doubles as a result but the (intensive) temperature stays the same. When someone fastens one steel bar to another just like it, the total (extensive) weight of the steel increases, but its overall (intensive) density is the same as before.

Probability is an intensive quantity because the likelihood of a particular possibility occurring in a random sequence is the proportion of the quantity of this particular possibility in the sample space to the quantity of alternative possibilities. If three out of the seven balls in an urn are red, the chances of pulling out a red ball at random are 3/7 (0.43). The probability of getting a 3 when throwing a die is 1/6, because there are six possible equiprobable events in that sample space. The probability of throwing at least one head in three tosses of a coin is 7/8 (0.875) since the sample space for the three tosses consists of eight possibilities and only one of these – TTT – is a head-free sequence. The probability of throwing at least two heads in three tosses is 0.5 because a half of the eight possible sequences in the sample space contains two or more heads (HHH, HHT, HTH, THH) and the other half contains two or more tails (TTT, TTH, THT, TTH).
Notice that the probabilities of the items in the sample spaces in these problems, and in every other probability problem, always add up to 1. Since all the sequences in three-coin tosses contain either two or more heads or two or more tails and the probability of getting at least two heads is 0.5, the probability of throwing at least two tails is also 0.5. The probability of throwing at least one head is 0.875, and so the probability of throwing three tails, and therefore no heads at all, is 0.125.

The crucial point here is that the calculation of the probability of an event or a class of events must be based on all the quantities in the sample space and not just the quantity of the event that we want to predict. I can only calculate how likely I am to draw a red ball from an urn that contains red, blue and white balls by working out the proportion of the red balls to the total number of balls in the urn (R/(R + B + W)). This might seem a simple and obvious point, but it is certainly not one that we can take for granted when we consider children's understanding of probability. There is evidence that children often restrict their analysis of the sample space to one quantity only. If they want to know the probability of pulling a blue ball out of the urn, many of them will attend to the number of blue balls but not to the number of the other balls in that container. This mistake is well documented and it represents a significant obstacle to children's thinking about probability. The challenge is to find a way of helping children to surmount this apparently formidable cognitive barrier. Fortunately, two factors should help us to do that. One is the high quality of the research on children's difficulties and successes in calculating and comparing probabilities. The other is the often neglected fact that children have exactly the same difficulties in calculating and comparing other intensive quantities, such as concentration of a solution and speed, as they do in dealing with probability. Research on how they learn about intensive quantities in general is a rich source for suggestions about how to help children to measure probability.

Calculating single probabilities

Infants

We start with problems in which children have to make some calculation about a single sample space. Most of these can only be solved by calculating a proportion, but there is a class of

extremely simple problems that are open to a simpler solution. Suppose that you were given an urn which contains eight red and four blue balls and are asked whether you would be more likely to pull out at random a red or a blue one. You could solve the problem by calculating that red is the answer because the probability of getting a red ball is 0.75 (8/12), but another easier way to reach the same conclusion is simply to register, without even counting, that there are more red than blue balls in the urn, and therefore a red choice is more probable than a blue one.

There are good reasons for wondering whether children can answer this kind of question in this kind of context. If they can, we can be reassured that they do have some knowledge that is relevant to probability. To realise that a red choice is more likely than a blue choice is to know in some sense that the probability of the two choices depends on the relation between the quantities of the two colours in the container. The relation in this case is a simple more–less relationship rather than a calculated proportion, but children who respond to this relation in a probability task would be showing at the very least that they are able to attend to more than one quantity in the sample space.

The evidence on children's success in this kind of task is encouraging. Several studies have shown that even babies of 12 months or less in age are apparently able to work out that the more numerous of two possibilities is the one more likely to occur. All these studies took advantage of the fact that babies tend to look for a longer time at objects and events that are in some way or other novel or surprising than at those which are familiar and to be expected, as we mentioned in the section on Randomness when we discussed the study by Xu and Denison (see page 28). The idea that these studies share is that if babies understand something about probability they will be more surprised by, and therefore will look longer at, a relatively improbable event than at a more probable one.

Teglas *et al.* (2007) showed 1-year-old babies a film of four objects whirling around in a container, which also had a pipe as an outlet. Three of these objects were in one colour and the fourth in another. At a certain point, the container was obscured, and at the same time the infants could see an object coming out through the exit pipe. Sometimes, this object was one of the three in the more frequent colour: at other times, it was in the unique colour. The experimenters reported that the infants looked for a longer time at the unique object exiting than they did when one of the three objects with the same colour as each other emerged. They argued that the babies had judged that one of the more frequent objects would be likely to exit, and were surprised by seeing the unique object come out because they knew this to be a relatively improbable event. It should be noted that, although the unique object was less likely to exit, this was by no means an impossible event. It is quite possible that perhaps adults would not be surprised at the 1-in-4 event taking place instead of an even more probable one.

However, the experimenters recognised that the infants' greater interest in the unique than in the frequent objects may not have been due to the differences in their probability. An alternative reason for this result could simply have been that the frequent objects were more familiar to them since there were more of them, and they therefore attended more to the relatively unfamiliar colour when that came out than at the colour that was more heavily represented in the container. So, Teglas *et al.* conducted a second experiment in which the container in the film was divided by a wall into two compartments. The wall made it impossible for the objects in one compartment to exit through the tube, whereas any object in the other compartment could escape in this way. In this new experiment the three identical objects were placed in the first of

these compartments and could not therefore escape, while the other compartment held the unique object, which was therefore the only object that the laws of physics would allow to come out through the tube. In the films that the children saw, the object that eventually exited through the tube was sometimes the unique object that had been in the compartment with the exit tube (a possible event) but at other times one of the frequent objects in the other compartment (an impossible event).

The question posed by the experimenters was whether the children would respond on the basis of possibility and impossibility (they would be surprised by the impossible event, and therefore would look longer if one of the frequent objects emerged, but not by the less frequent event, in which case they would look less if the unique object was the one to exit. The results followed the first of these two patterns: the infants looked longer at the impossible event than at the possible one. Frequency seemed to play no part in their reactions to these films. From this second result, the experimenters argued that the babies must have been responding to the different probabilities of the two possible events in their first experiment, and not just to their familiarity.

We have some doubts about this conclusion because the first experiment was about possible events with different probabilities while the second was about impossible versus possible events. An alternative experiment would involve two containers with a separating wall, as in the second experiment, and in each of the containers the probability would be the opposite of the other: three red balls and one white ball, for example, in the left side of the container, and three white balls and one red ball on the other side. The infants would be equally exposed to the colours. If each compartment had a pipe, drawing a red ball from the left pipe and a white ball from the right pipe would be the most expected events. Thus, a more likely event could be compared with a less likely one without any of the biases in the previous experiments.

Nevertheless, Teglas *et al.'s* conclusion is in many ways startling and provocative. It amounts to a claim that babies either start their lives with the ability to judge the relative probability of specific events or acquire this ability very early on in their lives and long before they could possibly put these expectations into words. Since the conclusion is based on ingenious research we must take it seriously but also critically. Before we embark on a more detailed analysis, however, we should like to turn to a second set of studies which also suggests that babies understand probability in quite a sophisticated way and use this understanding to make predictions about what will happen next.

This research was done by Xu and her colleagues, some of whose work we mentioned in Section 2: Randomness and its consequences. The babies in Xu and Garcia's research (2008) were only 8 months old, and these researchers also used looking time as a measure of surprise in their research. After giving each of the babies some experience with boxes that contained coloured balls, the experimenter then put another box, now closed, in front of him or her and went on to take out apparently at random (she closed her eyes) five balls from the box and to display them to the baby without at the time letting the baby see contents of the box. At this stage, therefore, the baby could see all five 'sample' balls, four of which were white and one red (or vice versa). Next, the experimenter did open the box and revealed that it held a total of 75 red and white balls. Seventy of these were in one colour and five in the other. Sometimes the majority colour in the sample and in the box were the same (e.g. four out of five sample balls were red and 70 out of the 75 balls in the box were also red): at other times the majority colour in the sample and in the box were different (e.g. four sample balls were white but only five of the

75 balls in the box were white). The experimenters called the first kind of event ‘probable’ and the second ‘improbable’ on the grounds that it is more likely that a random selection from a box with vastly more red than white balls will also contain more red balls than white ones.

The results of this first experiment were simple enough. The babies looked longer at the improbable event (one majority colour in the sample and another in the box) than at the probable one (the same majority colour in both). The experimenters claimed this as evidence that babies as young as 8 months understand the link between quantitative relations in the sample space and the probability of specific events, but of course they were concerned, as Teglas *et al.* had been in their own research, that the result could be dismissed as a response to the actual frequencies of the two colours rather than as the product of an analysis by the babies of the relative probability of specific events. So, as a control study, they did a closely similar experiment with another group of babies of the same age. This new experiment differed from the first in only one way. Instead of drawing the sample from the box, the experimenter now drew it from her own pocket. Thus, the balls in the sample had no physical or logical connection to those in the box, but the distribution of the colours was exactly the same in this experiment as in the previous experiment. In this experiment, the babies paid much the same amount of attention to the box in the two conditions. They looked about the same amount of time when the majority colours in the two containers matched as when they did not. The difference between the two experiments led Xu and Garcia to conclude that infants do form logical judgements about the probability of particular outcomes because they are apparently surprised when the box turns out to hold many more red than white balls after a random selection from the box has produced more white than red ones. ‘The present studies’ they argue ‘provide evidence that early in development infants are able to use a powerful statistical inference mechanism for inductive learning’ (p 5015).

Together, these two studies make a powerful case for the idea that very early in life children already have an impressive understanding of probability and use this understanding quite systematically to form expectations and to make predictions. This claim has serious educational connotations: if it is right, there should be ways of capitalising on this apparently well-formed knowledge when the time comes to teach children how to solve formal probability problems, and it should help teachers to know exactly what form this knowledge takes. Our own view, however, is that we need to know much more about the strengths and the limitations of very young children’s understanding of probability before we can apply any of this research to teaching.

One issue that we still need to settle to is whether babies can distinguish impossible from improbable events, even though this distinction is a basic and crucial part of any analysis of probability. We know from Teglas *et al.*’s experiments that babies are surprised by impossible events and by improbable events too, but neither these experiments nor those by Xu and her colleagues tell us whether or not babies grasp the fact that that improbable events, though surprising, can and do happen.

Another issue is about understanding probability as a proportion. As we have pointed out already, most probability problems can only be solved with the help of a proportional calculation, but there are just a few of the problems that can be solved on the basis of simple relations like ‘more’ and ‘less’ – i.e. of an additive rather than a proportional relation. Suppose that you particularly want a green marble, and you know that jar A contains more white marbles than green ones (W > G) while jar B holds more green than white marbles (W < G). You could decide

quite logically that you would be more likely to pull out a green marble at random from jar B than from jar A on this information alone. Thus, the problem can be solved on the basis of the simple relations ‘more’ and ‘less’. It is true that you could also solve it by counting all the marbles and then calculating and comparing the proportion of green marbles in the two jars, but that would be cumbersome and, in this case, unnecessary. The simple and direct solution of relying simply on the more/less relation differs from the usual proportional solution to probability problems because the judgement can be made on the basis of a perceptual difference between the two elements and this makes it possible to avoid having to calculate a proportion.

Several studies, not directly about probability, have shown that, in general, children are much more successful when it is possible to solve problems on the basis of a simple more/less relation than when some proportional reasoning is necessary. Spinillo and Bryant (1991, 1999) showed that children easily discriminate containers with bricks in two different colours when the more/less relation between the two colours in one box is the opposite of the same relation in the other (e.g. red>blue in one and red<blue in the other). In contrast, when the difference between the two colours in the boxes is in the same direction (e.g. red>blue in both boxes) but there is a higher proportion of red in one box than in the other, the same children make many more mistakes.

Both the research teams whose work on infants we discussed chose problems that could be solved by direct more/less additive comparisons. There was no need in these problems for proportional comparisons. In both sets of experiments, the babies could have reacted as they did to improbable events on the basis of simple ‘more’ and ‘less’ relations between the two categories of elements in their experiments.

In the Teglas *et al.* study the babies only needed to know that there were more, say, blue than yellow balls in the container to be surprised that the first ball to emerge was yellow. They did not have to calculate the proportion of blue balls to the total number of balls to form the expectation that the first ball out would be a blue one. When, in the Xu and Garcia experiment, the babies first saw a sample of four red balls and one white ball (R > W), they could have formed the quite rational expectation that the box from which the sample came also would also contain more red than white balls. When the box was opened and they saw that it contained many more white than red balls, that expectation was violated. They expected R > W but they saw R < W, and they were surprised. Their reactions could have been based entirely on their expectations about simple ‘more’ and ‘less’ relations between the two parts (colours). In neither of the two studies do the researchers go into any detail about the type of calculation that that they think the babies made, or about the type of relation that they think the babies used to make their calculations, though a footnote by Xu and Garcia implies that they think that these calculations were a great deal more sophisticated than the ones that we have just suggested.

Thus, the two studies leave us with a serious question: Are babies of this age able to form expectations on the basis of proportions or are these expectations based entirely on simple ‘more’ or ‘less’ relations, as we have suggested? The studies also provide us with ideas about how to answer this question. The methods that these researchers have pioneered so successfully can easily be adapted to deal with this issue.

Here is a new version of the Xu and Garcia method that would effectively tackle the problem. Show the babies a sample of six red balls and then reveal a box that contains on different occasions:

a. 65 red and 10 white

b. 40 red and 35 white

c. 40 white and 35 red

d. 65 white and 10 red balls.

The simple more/less relation in the sample is R > W and this relation also characterises the contents of boxes (a) and (b) but not (c) and (d). So, if we are right in thinking that babies respond only to simple relations, they should be surprised by the contents of boxes (c) and (d) in which the opposite relation holds but not by boxes (a) and (b). However, for proportional reasons, a random selection from box (a) would be far more likely to produce a sample of six red balls and one white one than a random selection from box (b). So, if babies do process proportions they should be more surprised at the contents of box (b) than at those of box (a) and should spend more time looking at box (b) than at box (a). This seems an unlikely result to us but it could happen and, if it did happen, it would have extremely important implications, for it would mean that children start their lives in possession of a sophisticated mechanism to help them learn about and understand probability. We have to wait for the answer.

Primary school children

It is not a big jump from the possibility (but improbability) that babies calculate proportions to the question of how children aged 10 years or more learn how to give a proportional number to probability when they are taught about this in school. The work on this topic has produced two main results. The first is that initially children find it hard to understand and apply what they are taught about proportional calculations of probability. The second is that children's success at applying what they have been taught to the calculation of probabilities improves rather sharply between the ages of 10 and 13 years.

The classic study on this issue was done by Fischbein and Gazit (1984) with 10- 11- and 12-year-old students who had been through a 12-lesson course on the basic concepts of probability, which included instruction on how to make the correct proportional calculation in a range of contexts. The researchers then assessed the children's understanding of several aspects of probability, one of which was how well they were able to work out the probability of various events. For example, in one set of questions they were told first about a box that contained four black and three red marbles, and then were asked to work out the probability of:

a. picking a red marble at random from this box (3/7, 0.43)

b. picking a black marble (4/7, 0.57)

c. after picking and then replacing a black marble, picking another black one the next time round (0.57)

d. after picking a black marble but not replacing it, picking a black one the next time round (3/6, 0.5).

Notice that the first three questions are about independent events whereas in the fourth the sample space changed as a result of what happened in the first draw.

The 10-year-old children made many mistakes in their answers to all four questions. Their best score was 30.9% correct answers to the second question, (b): their rate of success with the other three questions was well below 30%. The 11-year-olds did a great deal better in the first three questions which were about independent events, but rather badly (well below 50%) with the fourth question which required them to recalculate the sample space before doing the proportional calculation. The 12-year-olds' also did better in their answers to the first three questions than to the fourth question, but their scores were all quite high. Their answers were correct over 85% of the time to the first three questions, and 71% of the time to the fourth question.

The most dramatic feature of these results is how much more successful the older students were than the younger ones at calculating quite simple probabilities. This cannot have been due just to teaching about probability, since all the children who were asked these questions, according to the authors, had been given the same amount of teaching on this subject. However, it may have been due to the older students having had more teaching and a great deal more experience than the younger students in dealing with fractions as well as with proportional problems both in their mathematics and in their science lessons. It would be wrong to think that the only teaching given to these children that would help them calculate probabilities was their 12-lesson course in probability.

The students' commonest mistake in the answers to the fourth question was to forget to adjust the number of the total (from 7 to 6) and the number of black marbles (from 4 to 3). In the students' answers to the first three questions, one of their most frequent mistakes was to try to carry out an additive rather than a proportional comparison. The students often gave the frequencies of the different marbles as the probability of drawing a red marble at random but, as they had learned to express probabilities in fractions, they answered that the probability of drawing a red marble was 3/4 and that the probability of drawing a black one was 4/3. If they had understood the connection between ratios and proportions, they would have realised that 4/3 represents a quantity larger than one, and is an impossible answer in probabilities. This is an interesting and important observation, since it suggests that schoolchildren may be readier to compare ratios than fractions, as long as they understand that comparing ratios is not about a simple subtraction between the frequencies of events, and therefore takes us back to our discussion of the research on infants which featured probability tasks that could be solved through comparisons between parts. The children's mistaken attempt to base their calculation on the two parts points to the possibility that the step from part–part additive comparisons to ratio or part–whole, fractional comparisons in all relational problems, including of course probability ones, is a genuinely difficult one for children to take.

Comparing two (or more) probabilities

So important is proportional reasoning in calculations about probability that much of the research that has been done on children's understanding of probability is about their ability to understand and manipulate proportions and nothing else. This widespread interest in the link between understanding proportions and understanding probability is justified, since, as we have already seen and will continue to see in the rest of this section, the proportional nature of

probability is a source of genuine difficulty to many young children and to older school students as well.

The difficulties that children have with the proportional element in probability problems was a central theme in Piaget and Inhelder's (1975) book on children's understanding of chance, and their powerful experimental research on children's comparisons of different probabilities is a good starting point for a discussion of children's proportional reasoning about probability. The question that these authors posed was how children set about comparing the probability of a particular event in two different sample spaces. The task, which they gave to children whose ages ranged from 6 to 13 years, was simple but effective.

The researchers used counters, some of which were plain on both sides while others had a cross drawn on one side and were blank on the other. Over several trials, the researchers started by forming two sets of these counters. First they showed the child the contents of each set and in particular how many counters in each set had crosses on them and how many were completely blank. Next, they turned the counters with crosses over so that the child could only see their blank side, and shuffled the positions of the counters in each set. Finally they told the child to take out one counter at random, but to draw it from the set from which she would be most likely to pick a counter with a cross. They also asked the children to give their reasons for choice of set.

In some of the trials there was no uncertainly about what would happen if the child chose a counter from one at least of the two sets. Either all the counters in the sets had a cross on them or none of them did. In other trials, there was a mixture of blank and crossed counters in both sets. These are the trials that we shall concentrate on at this point because they are most relevant to the question of how well children reason about proportions and to what extent they resort to analysing additive relations instead.

When both sets were a mixture of crosses and blank counters, the right set to choose was the one with a higher proportion of crosses. However, in some trials there were alternative solutions. For example, in one trial the two sets contained the same total number of counters in each, but an unequal number of crosses (e.g. one set had 1 crossed and 3 blank counters and the other 2 crossed and 2 blank counters): in these comparisons, the child could avoid carrying out a proportional analysis and could solve the problem just by directly comparing the number of crosses in the two sets (1 cross in one set but 2 crosses in the other). In other trials, the number of crossed counters in the two sets was the same, but one set contained more blank counters than the other. In these trials, once again, the children could solve the problem without carrying out a proportional calculation: the more blank counters there are, the smaller the chance of drawing a crossed one. Piaget and Inhelder called these problems 'one-variable problems', because they did not require a comparison of the ratio or proportion of positive versus negative cases across the sets.

In other trials these simple types of solution were not possible, because both the number of crosses and the total number of counters differed between the two sets. For example, one set contained 1 crossed and 2 blank counters, and the other 2 crossed and 4 blank counters. Here, Piaget and Inhelder argued, the child had to work out the ratio (in both cases, there is a 1:2 ratio of crossed to blank counters) or the proportion (one third of the counters are crossed) of crosses in each set in order to solve the problem.

Piaget and Inhelder's account of the results of this study dwells on children's justifications for the choices that they made. The youngest children made very little attempt at a quantitative analysis, and when they did quantify anything it tended to be a single quantity (usually the number of crosses) in each set. Slightly older children (about 7 years or so) were more systematic and more successful, but their successes came mainly in trials in which they did not have to carry out a full proportional analysis of the contents of each set, as in the one variable problems. This was shown in problems in which the number of crosses and the total number of counters were both different in the two sets. Thus, some kind of a proportional comparison was necessary in these problems, and these children usually produced the wrong answer, which they often justified by arguing about one of the parts – usually the absolute number of crosses – in each set without relating this either to the total number of counters or to the number of blank counters.

The oldest children tended to use the ratios in their comparisons; from about age 9, they could, for example, with the problem that we have just mentioned argue that in each set there were two blank counters for each crossed one. Piaget and Inhelder described this strategy as 'the construction of empirical ratios'. They claimed that it represented a change from additive to proportional reasoning in the children's solutions. Piaget and Inhelder also noted that these ratio comparisons were more effectively carried out when the ratios in the two sets were the same. When the children explained their reasoning, they used phrases such as there are three times as many crosses in both sets or indicated that the proportion was the same (there are two and two here, and one and one here).

The comparisons that the children based on empirical ratios were not as likely to lead to the right conclusion when the probabilities were unequal. For example, an 11-year-old child, who had succeeded with sets with equal probabilities, did not manage to make a successful comparison between a set in which 1/3 of the counters had crosses and another set in which 2/5 had crosses. After setting the counters in 1 to 2 correspondence, the child concluded that the probabilities were the same because 'there is one fewer without a cross in what remains' (p 155). Similarly, a 12-year-old who correctly solved the comparison between two sets in which there was the same proportion of crossed to the total number of counters (1/2 and 2/4) concluded that 2/5 and 6/13 will have the same probability of drawing a counter with a cross: 'There is as much risk on one side as on the other. Here there are six with crosses and seven without crosses, and there two and three [a difference of one on each side]' (p 156)

Finally, many of the oldest children, usually children aged 10 years or more, solved the more difficult problems, like the comparison between a set of 2 crosses and blank counters versus a set of 3 crosses and 6 blank ones. These children also consistently referred to the proportion of crosses to the total number of counters, when justifying their responses, or to the ratio between crossed and blank counters in each set.

Piaget and Inhelder argued that the three patterns of responses correspond to three, successive developmental stages. In the first, children have no consistent, recognisable cognitive approach to the quantification of proportional problems, although they distinguish possibility from impossibility. In the second, they solve problems which do not need a full proportional analysis, and do well where it is possible to solve the problem by a direct comparison between a single part in one set with a single part in the other set (e.g. directly comparing the number of crosses or the number of blanks in the two sets). The third stage is the

final one: the children who reach it do, according to Piaget and Inhelder, carry out a full proportional analysis of each set.

Piaget and Inhelder's method for interviewing the children, the clinical method, does not involve simply presenting the children with questions to be answered. Using the clinical method, Piaget and Inhelder often confronted the children with alternative ways of reasoning about the same problem: for example, if a child said that it would be best to choose a counter from a set with 2/6 crossed counters than from another with 1/3 crossed counters, the experimenters would rearrange the counters, setting them in 1 to 2 correspondences, and asked whether the child still thought that it was best to choose from that set. It is thus by no means certain that these empirical correspondences would be used by children answering tests without such counter-suggestions of different ways of thinking about the problem. The clinical method here could have worked almost as a teaching experiment, in which the teacher provokes conflict between different ways of approaching the problem without telling the solution. This can be seen as both a strength and a weakness in their approach.

Piaget and Inhelder might have obtained the best performance from children who reacted positively to such counter-arguments but it is unclear whether children would arrive at this reasoning without their prompts.

This is an ingenious, and justly famous, experiment. It provoked a great deal of interest and led, as we shall see, to many other rather similar studies by other research teams. How justified are the conclusions that Piaget and Inhelder drew from it? Broadly speaking, their suggestion, that initially young children do not compare sample spaces on the basis of proportions and concentrate instead on single parts (e.g. the number of crosses) in the two sample spaces that they are comparing, seems unassailable. As we shall see, subsequent research has consistently confirmed that the genuinely proportional problems that Piaget and Inhleder devised are indeed far too difficult for most children under the age of 10 years.

These researchers' other main conclusion was that time (and age) take care of this difficulty with proportions, so that eventually all, or very nearly all, children will be able to solve even hardest of the problems. This is more questionable. For one thing, as we have already mentioned, there is evidence that most 15-year-old students – students, that is, who were older than those who took part in Piaget and Inhelder's study – cannot solve problems that are very like the ones that we have been discussing. As we have already mentioned, 73% of a large group of European 15-year-olds failed to make the right choice when they were given the following choice in a PISA test (Pisa Consortium Deutschland, 2004)): 'Box A contains 3 marbles of which 1 is white and 2 are black. Box B contains 7 marbles of which 2 are white and 5 black. You have to draw a marble from one of the boxes with your eyes covered. From which box should you draw if you want a white marble?'

This astoundingly high rate of failures seems all the more remarkable when one considers that all the students had to do was to pick the right box out of two boxes, and thus, 50% of them could have been right by chance alone. Yet their success rate was far lower even than that. It is possible that the students were led astray by a direct comparison between the number of white marbles across the sets without considering the proportions. The problem does not actually require difficult calculations and could be solved by ratio reasoning: there are twice as many white marbles in the second box but more than twice as many black marbles in the second box, which makes its choice less advantageous. This generally unimpressive performance does not

seem to us to be consistent with the claim that eventually all the students will be able to compare sample spaces proportionally.

An alternative hypothesis, which Piaget and Inhelder's account of their own research certainly does not rule out, is that some children, more mathematically gifted than others, learn about the importance of proportions in calculating probability either as a result of their informal experiences or through being taught about probability, while others do not. We badly need longitudinal research which traces students' reasoning about comparable sample spaces right through adolescence and into adulthood.

Piaget and Inhelder's description of the children's reasoning raises another issue which may seem at first to be merely a technical one, but is, we think, of enormous significance in the study of children's learning about probability and about proportions in general. There are two ways to calculate a proportional relation. One is to calculate the relationship of one part to a whole, a relationship which is usually represented by a proportion and thus a number smaller than 1. So, when 2 out of 6 counters have crosses on them, the proportion of crosses is 2/6 or 1/3 or 0.33. The other is to calculate the relationship between the two parts, and to form a ratio between them: in the same sample space that we started with, the two parts are crossed and blank counters, and the ratio of crossed to blank counters would be 2:4 or 1:2, which is another way of saying that for every one crossed counter there are two blank ones in the sample space. Ratios can be represented by whole numbers, and children are considerably more familiar with whole than with rational numbers. All of Piaget and Inhelder's quantification of probability problems could be solved in either of these two very different ways, and in fact many of the comments made by the most successful children in their study suggest that they were making ratio judgements and not part–whole ones, that would involve fractional representation. For, example, here is the reasoning of a 12-year-old student who is comparing a set (A) of 1 crossed and 2 blank counters to a set (B) of 2 crossed and 3 blank counters: 'It's easier here (Set B) because it gives you 2 chances against 3, and there (A) 1 against 2'. This seems to us to be a comparison of the ratios between crossed and uncrossed counters in the two sets, and so do the remarks of a 10-year-old who was asked to compare one set with a single crossed and a single blank counter to another set with 2 crossed and 2 blank counters: 'There are the same number with crosses and without crosses in both groups.' In both examples, the students are clearly reasoning about the relationship between the two parts of each set and appear to be forming a ratio between them on the basis of one-to-many correspondence (there is 1 chance against 2) or many-to-many correspondence (there are 2 chances against 3).

The distinction between part–whole proportional solutions and solutions based on ratios formed through one-to-many and many-to-many correspondence is highly relevant to educational issues. When there are two effective and intellectually respectable ways of solving the same problems, we need to wonder which way is most easily learned and which leads to the greater progress in the pupils' understanding, in this case of probability, in the long run. This is surely a matter for teaching experiments and for longitudinal research.

Now, we turn to the research on children's comparisons of probability which followed, and was certainly provoked by, Piaget and Inhelder's study. As we have already mentioned, there is ample evidence of the same difficulties with proportional reasoning in this subsequent research. The well-known study by Fischbein and Gazit (1984), a part of which we described earlier in this section, also included a task in which children had to compare two probabilities, which were equal. 'Uri has in his box 10 white marbles and 20 black ones. Guy has in his box 30 white and

60 black ones. The winner is the child who pulls out a white marble first. Uri complains that the game is not fair because there are more white marbles in Guy's box than in his box. What is your opinion about this?' This problem is equivalent in the demands that it makes on proportional reasoning to one of Piaget and Inhelder's more difficult problems, and it did prove difficult, particularly for young children. The success rate for children who had recently taken a 12-session probability course was only 21.1% for 10-year-olds, 53% for 11-year-olds and 69.2% for 12-year-olds. These children did no better, and in fact in one age group did rather worse, than other students of the same age who had not been taught about probability. The children's answers therefore throw some doubt on the value of formal instruction in this area, at any rate on the value of the type of formal instruction that they had received, and at the same time confirm both the difficulties that young children have with this kind of problem and also the claim that these difficulties decline with age. The fact that the 12-year-olds appear to have been more successful at comparing probabilities than those 15-year-olds who were given the PISA task is a bit puzzling, but it may be due to Fischbein's comparisons being between equal probabilities and the PISA problem about unequal ones. Equal probabilities may be easier to detect using ratio comparisons, as suggested by Piaget and Inhelder. One interesting post-script to the Fischbein and Gazit results is that they mention that nearly all the 12-year-olds and over half of the 11-year-olds who solved the problem also referred explicitly to the white–black ratio in the two sample spaces in their answers. This is further evidence that the use of ratios may be the natural way for children to solve probability problems.

The methods that Falk, Falk and Levin (1980) used to study quite young children's comparisons of different probabilities were based on games of chance, appropriately enough because the serious study of chance and probability seems historically to have been largely provoked by questions about gambling (Mlodinow, 2009). These experimenters presented two studies, only one of which we shall describe here. In this, the children, whose ages ranged from 4 to 7 years, were given two roulette-type wheels of different sizes, which were divided into blue and yellow sections. The relative amount of blue and white differed between the two wheels, and so did the number of sections on the wheel: these two variables were confounded: the larger of the wheels contained more sections than the smaller one. The child's task was to pick the wheel that was more likely to stop at blue (or yellow). Since the actual areas and the number of elements in the two wheels were very different from each other, the correct solution to the problem had to be based on the relative size of the blue and white areas, or on the relation of the blue area to the total area, within each wheel.

Given the other studies that we have reviewed, the performance of the young children in the Falk *et al.* study was very good. Nevertheless, Falk *et al*. show that the younger children in particular tended to be thrown of course by the relative number of elements of the chosen colour in the two wheels. They tended to pick the wheel with the larger number.

Finally, we will describe a study by Falk and Wilkening (1998) which was also about children's comparisons of probability, but which used a different and promising new technique. In their task, which they called an 'adjustment task', they presented to children whose ages ranged from 6 to 13 years one urn that contained beads of two different colours and another urn that had beads in one of the two colours only. They explained that the urns would be part of a lottery game, and that one of them (the experimenter and the child) would draw a bead from one of the urns and one from the other. The winner would be the one who drew a bead of a particular colour, but first they had to ensure that they both had the same chance of drawing a bead of that colour. The experimenter then asked the child to complete the contents of the second urn

by putting in beads of the missing colour, in such a way that each of them would have the same chance of getting the winning colour, and added that s/he (the experimenter) would choose which urn she used and which urn the child used in the draw. Each child was given several different problems of this sort, and the proportions of the winning and losing numbers as well as the absolute numbers varied from trial to trial.

The main advantage of this adjustment method, apart from being an interesting way of recruiting the children's attention and interest, is that the mistakes that the children made should tell us something about the strategies that they were using. For example, some children might match the two sample spaces by putting an equal number of the winning colour in both urns. Others might respond proportionally, by putting in, for example, twice as many winning as losing balls in the incomplete urn when this was the relationship between the two colours in the already completed urn.

In the event, in a rather sophisticated analysis of their results, the researchers detected three clear patterns among the individual children. Some children's adjustments reflected the number of one of the two colours only in the already competed urn: for example, the more winning balls in the complete urn, the more of the same coloured balls they put into the urn whose contents they had to complete themselves. Others adopted an apparently more complex strategy, by taking into account of the number of both of the colours but in an additive, not a proportional way. So, the number of winning colour balls that they added was determined by how many balls of the same colour in the other urn and also by the absolute difference between the number of winning and losing colour balls in that urn. Falk and Wilkening called this pattern the 'difference' strategy. The third and successful strategy was proportional. The children's adjustments to the second urn reflected the ratios of the two colours in the first urn. If there were three times as many winning as losing balls in one urn, they reproduced that ratio in the second urn as well. The equal ratios ensured that the proportions in the two urns were the same.

There was, as one might expect, a strong relation between the children's ages and which of these three strategies they tended to adopt. Those who attended to the quantity of one of the colours only were younger on average than the children who apparently took both colours into account, but in an additive way (the difference strategy), and these children tended to be younger than the children who adopted a proportional strategy. This last strategy was the rarest, and most of the children who adopted it were in the two oldest age groups, the 11- and the 13-year-olds. Hardly any children younger than that carried out the adjustment of the contents of the second urn proportionally. Thus, there is a mismatch between the results of earlier Falk *et al.* study and the more recent research by Falk and Wilkening; many 6- and 7-year-old children apparently did use proportions to compare two probabilities in the Falk *et al*. study and yet it is hard to find any sign of proportional reasoning in children under the age of 11 years in the Falk and Wilkening adjustment experiment.

Falk and Wilkening comment on the apparent difference between their results and those in the Falk *et al.* study and other similar investigations into children's understanding of probability. They offer two possible explanations for the discrepancy. One is that it may be harder to adjust quantities proportionally than it is to compare them proportionally. This is speculation, but it might be right. The other is that some of the children's successes in the earlier research on children's comparisons of two probabilities may have been due to children using a version of the difference strategy, which they spotted among many of the children in their own study, rather than a genuinely proportional strategy to make their comparisons. This does not seem to us to

be a convincing analysis. It does nothing, in our view, to explain how well the children did in the Falk *et al*. study when for example both wheels contained a larger blue area than a white one, but the proportions of the two colours were different in the two wheels.

Summary of the work on comparisons of probability

The remarkable research on babies and probability suggests that, early in their lives, children grasp the relationship between the relative number of possible events and the likelihood of particular events occurring. The infants seem to realise that one is more likely to pull a red than a blue ball out of an urn that contains more red than blue balls. However, when infants and children up to age of around 10 years have to go beyond simple more/less relations to calculate the probability of particular events, they usually encounter a great deal of difficulty. The obstacle, which is also an obstacle to learning in several other parts of young children's intellectual life, is the need for proportional reasoning in most probability problems.

Research on children's attempts to compare probabilities across two sample spaces has established the importance of the distinction between using difference relations and making proportional comparisons to solve probability problems. The most difficult of these comparisons are undoubtedly those that demand proportional reasoning. There are some rather surprising discrepancies between the data that we have from different studies on how difficult these proportional comparisons are, but it is clear that problems, that can only be solved on the basis of proportional reasoning, are much harder for children up to the age of roughly 10 years than problems that can be solved in some other way.

Research on this cognitive demand has concentrated on how serious an obstacle it is, and has neglected the different ways in which children tackle proportional problems. Yet, details in the results of several different studies point to a distinction of great importance to the study not just of children's cognitive strategies but also of ways of teaching them about proportions in general and probability in particular. This is the distinction between fractional and ratio representation in relational comparisons.

Fractional representation, which is the conventional basis for calculations about probability, deals with the relation between part of the sample space and the whole sample space. If there are 6 red balls in an urn and 3 blue ones, the probability of pulling out a red ball is 6 divided by the total number of balls, 9, which comes to 0.67 or 2/3. Ratio representation deals with the part–part relation in a proportional way: in the same urn, there is a 2 to 1 ratio of red balls to blue balls.

In this section, we have encountered evidence for two kinds of part–part reasoning. One is part–part reasoning that deals only with difference, which is a form of additive comparison between the parts, and is not genuinely proportional. So, some children can discriminate two containers by the pattern of their contents, if one clearly contains more red than blue marbles while the other contains more blue than red. They understand the difference between R > B and B < R, and use it to solve apparently proportional problems. This relational, but not proportional, solution is very common among young children's judgments about area and number (Spinillo and Bryant, 1991, 1999), and applies to probability as well, as shown in Piaget and Inhelder study.

The second form of part–part reasoning is based on ratio, and genuinely proportional; we argue that it does underlie many of the older children's successes in the comparison tasks, for example, in Piaget and Inhelder's and in Fischbein's research. This is children's calculation of the ratios between the different elements of the sample space on the basis of one-to-many and many-to-many correspondences. In the next section, we briefly review research that shows that the question of how multiplicative relations are represented definitely has an impact on children's problem solving and learning.

Teaching children about proportional calculations

Ratio and fraction language in other proportional problems

We started the analysis of quantification of probabilities by referring to the distinction between extensive and intensive quantities; the latter always involve proportional reasoning. Some intensive quantities, and this includes probabilities, can be meaningfully represented by either fractions or ratios: a ratio of 1 to 2 crossed to blank counters indicates exactly the same thing as 1/3 crossed counters or a probability of 0.33 of drawing a crossed card. Other intensive quantities that can be expressed as ratios or fractions are, for example, the relative concentration of two liquids in a mixture (e.g. orange concentrate and water) or the density of objects in an area (e.g. flowers in a flower bed). On the basis of the studies reviewed in the previous section, and in particular the study by Piaget and Inhelder, we hypothesised some time ago (Nunes and Bryant, 1996) that children would be able to make more sense of intensive quantities if they were presented with problems in ratio rather than in fractional language. We know only a handful of studies that have analysed if presenting these problems to children using ratio or fraction language affects their problem solving performance or their learning.

Desli (1994) presented intensive quantities problems about the concentration of mixtures of liquids either using a ratio or a fractional representation in the problem presentation. For example, in the ratio language, the children were told that a child made orange squash using 1 cup of concentrate and 2 cups of water and found it tasted perfect; on another day, the child needed to make a much larger amount of juice (or a much smaller amount) because lots of friends were coming; she needed to make 18 cups of orange juice; how much concentrate and how much water should the child use? The same story was presented with fraction language: the perfect mixture was described as having 1/3 concentrate and 2/3 water. Other mixtures (tins of white and blue paint, for example) provided a variation in the context. The children were in the age range 8 to 10 years and attending schools in London. They had been taught considerably more about fractions than about ratios in school, as ratios are not an important part of the curriculum until about age 10. The study was a within-participants study, and each child solved half of the problems presented in ratio and the other half presented in fraction language. The language of problems was changed across participants: half of the participants answered particular problems in ratio language whereas the other half answered them in fraction language. For the 8- and 9-year-olds, there was a striking and statistically significant difference in the rate of correct responses.

As **Table 2** shows, both groups solved many more of the problems presented in ratio language than of those presented in fractional language. In contract, the 10-year-old children did equally well with both kinds of problem. The earlier success in problems presented in ratio language is in line with the results observed by Piaget and Inhelder, who showed that children in the age 9 to 11 years were able to construct empirical ratios to solve probability problems and used ratio

language (three times as many crossed counters than blank counters) when justifying their responses, before they could solve the problems by using part–whole quantification of probabilities.

Table 2: *The relative success of 8- to 10-year-old children with problems presented in ratio and fraction language*

Age	Correct response to problem in ratio language	Correct response to problem in fractional language
8	37%	12%
9	54%	23%
10	67%	65%

Subsequently, we (Nunes, Bryant, and Hurry, 2004) assessed whether children benefited as much from teaching about intensive quantities when the language used during the teaching experiment was either entirely in ratios or in fractions. The study used a pre-test, immediate post-test, and delayed post-test design; the teaching intervention was carried out after the pre-test and before the immediate post-test. The pre- and post-tests were identical; the children were asked to make relational comparisons that we expected could be solved without calculation and also to enlarge quantities while keeping their quality (taste, colour, density) constant. The children (N = 132) came from three different schools in Oxford and were in the age range 7 to 8 years. They were randomly assigned to one of three teaching groups: intensive quantities using ratio language, the Ratio Group (N = 46); intensive quantities using fractions language, the Fractions Group (N = 46); or computations with extensive quantities, the Control Group (N = 42). The random allocations were restricted so that approximately equal numbers would be assigned to each group in each participating school; because the teaching was carried out in pairs, the random allocation of two extra pairs to each of the intervention groups in two of the schools produced a different total number of participants across groups. The pairs of children were taught by a researcher, outside the classroom. The problems presented to the two groups taught about intensive quantities were the same, just the language of presentation differed. The problems presented to the control group, who worked with extensive quantities, involved multiplication and division, as did those presented to the intensive quantities groups.

This study produced three main results.

1. The pre-test scores were rather low: children in this age range showed little insight into the relational reasoning and the calculations necessary to enlarge an intensive quantity while preserving its quality (the same colour, the same taste).

2. The training had a clear effect. The children in the two groups taught about intensive quantities made more progress from pre-test to immediate post-test in their total scores for the intensive quantity problems than the children in the control group. This difference was still significant at the delayed post-test two months later. Thus the experiment establishes that it is possible and worthwhile to teach young school children about intensive quantities.

3. The Ratio Group improved more than the Fractions Group in solving the problems in which they had to make numerical calculations. There were no differences in the problems in which relational reasoning without calculations was sufficient to solve the problem.

This result was replicated with a very large sample (535 children from 24 classes) of Scottish children in the age range 9 to 11 years (Howe, Nunes and Bryant, 2010). In this study, the children were taught in groups by a researcher but the teaching took place in the classroom. The Ratio Group produced significantly more correct solutions and explanations for their solutions than the other two groups in the immediate and delayed post-tests. The Fractions Group, however, improved significantly in their ability to represent intensive quantities using fractions, and obviously did so significantly better than the Ratio Group, who had not used fractions language during the teaching. Surprisingly, though, the Ratio Group improved in the use of fractions language and at the delayed post-test caught up with the fractions group and significantly outperformed the control group.

The results of analyses of how children solve probabilities problems led us some years ago to raise the hypothesis that ratio language is an easier representation for children to learn, as they seemed to express their reasoning in ratio terms earlier than in fractional terms in the study by Piaget and Inhelder. We tested this hypothesis first in a problem solving context with intensive quantities and subsequently in two teaching studies; none of these studies actually included probability problems but focused instead on other intensive quantities that we thought might be more familiar to children. These studies converge in supporting the idea that children can more easily think about intensive quantities using ratio than using fractional representations. It remains to be seen if the same holds true for solving probability problems and learning about probabilities. There is some research that suggests that probability problems are better understood if presented in ratios than in percentages or proportions, which we review in the next section, but we know of no relevant teaching studies so far.

Teaching children to calculate probabilities

We have already made the argument that it is possible to teach children about intensive quantities, and we have also claimed that the most effective way to start this teaching is to use ratio language. These ideas have not been applied yet to children's ideas about probability, but we believe that this would be a promising way to approach the question of how to teach children to calculate and to compare probabilities.

The outcomes of previous intervention studies on these skills are mixed. Two groups took part in Fischbein and Gazit's study, which we have mentioned several times already. The children in one group went through a course on probability at school while the children in the other group were not taught about the subject at all. The two groups of children were then given a questionnaire that included items on comparing probabilities. The children's success in answering these questions suggested that the teaching given to the first group of children had had little effect on their ability to think about probability quantitatively. In fact, slightly more children in the control group than in the intervention managed to solve these comparison problems. However, it is difficult to be sure about what this result means because the researchers did not include a pre-test. Perhaps, by chance the control group children understood the basic rules of probability at the start of the study as well as at the end.

A study by Castro (1998) on teaching Spanish children to calculate, and to reason about, probability did include a pre-test as well as a post-test, and Castro also made a comparison between two groups of children. The children in one group were taught about probability in what Castro called the 'traditional' way, whereas those in the other were taught through methods that concentrated on, 'conceptual change'. The traditional methods apparently consisted mainly of the teacher telling the children how to carry out the correct arithmetical procedures. The method called 'conceptual change' involved encouraging each of the students to come up with their own ideas for solving probability problems, which were then discussed in the classroom by the other students. This is the most careful and the most effectively designed study of how to teach children about probability that we know of, and fortunately its results were positive. The children in two groups reasoned about probability and calculated probabilities about as well as each other, but by the time the intervention study was over the children in the conceptual change group were far ahead of the traditionally taught children in reasoning and in calculating probabilities. These striking results suggest that open classroom discussions are an excellent and interesting way of introducing children to probability, and of helping them to learn how to calculate probabilities.

Conditional probabilities and Bayesian reasoning

Often, the probability of one event depends on the probability of another. Take as an example a problem which Tversky and Kahneman (1982) devised and gave to adult participants, probably to the dismay of these participants because many of their answers were quite wide of the mark: 'If a test to detect a disease whose prevalence is 1/1000 has a false positive rate of 5%, what is the chance that a person found to have a positive result actually has the disease, assuming you know nothing about the person's symptoms or signs'. The question is about a 'conditional probability' because your confidence that a positive result indicates the presence of the disease should depend not just on the accuracy of the test itself but also on the number of people in the population who suffer from the disease in question. In this example, it is reasonably easy to see that the answer should be that the probability that someone who tests positive actually has the disease is rather small: in a population of 1000, we would expect the test to produce positive results, with 1 person who actually has the disease and with 50 healthy who do not. So, only 2% or thereabouts of the people who test positive could be expected to have the disease.

Many of the adults who were given this or similar questions answered quite differently. They maintained that the probability was 95% or 0.95 (rather than the much more appropriate 2% or 0.02), which indicates that they took into account the information about the accuracy of the test but not about the disease's prevalence. Yet, prevalence figures should be an essential part of the solution to a problem of this sort. To see this, one only has to think what the answer would be if the prevalence rate for the disease was 50/1000. In this case, about half the people who tested positive would be expected to have the disease.

It may seem perverse of us, having spent so much time discussing the difficulties that children have both in understanding and calculating straightforward probabilities, to inject into the discussion a new kind of probability problem which is clearly far more complex than the ones that have come before. We have two reasons for doing so. The first is that in the modern world we increasingly need to be able to think our way through conditional probabilities. To interpret much of the information that we are given about the probabilities of particular events and outcomes, we have to take account the probabilities of other events as well (as the Tversky–Kahneman problem neatly illustrates), and it will surely benefit the children in our society to

prepare them for this kind problem. The second is that there is some evidence that many quite young pupils can, in some circumstances but not in others, solve these problems remarkably well.

Recently there have been several reports (Gopnik, *et al.,* 2004: Sobel, Tenenbaum and Gopnik, 2004; Sobel and Kirkham, 2007; Xu and Tenenbaum, 2007) about pre-school children's ability to solve conditional problems using a form of Bayesian-reasoning, which, as we shall soon show, is often applied to conditional probability problems, but this research was about children's understanding deterministic cause-and-effect chains, and so we shall not describe it here. There is some direct research on children's ability to solve conditional probability problems, though not a great deal of it. The most interesting study on children's solutions to conditional probability problems, it seems to us, was done by Zhu and Gigerenzer (2006). These two researchers were interested in the possibility that children's success in solving conditional probability problems depends very heavily on the way these problems are presented to them. They argue that if the relevant information in the problem is provided to them in the form of probabilities, children without exception flounder. If, on the other hand, this information takes the form of frequencies, they often find the right solution and their rate of success improves with age. Their evidence for this claim is a study with Chinese children at school in Beijing; their ages ranged from 9- to 11 years.

The researcher worked with a set of ten problems, which they presented in two different ways. One way was to give the information as proportional probabilities in percentage form. The other way was to give this information in the form of 'natural frequencies': in this case, their stories were all about the absolute numbers and not about proportions. Here are two examples of the same story being presented first with proportions in the first version and next with frequencies.

A: 'Pingping goes to a village to ask for directions. In this village the probability that any person he meets will lie is 10%. If a person lies, the probability that he has a red nose is 80%. If a person doesn't lie, the probability that he has a red nose is 10%. Imagine that Pingping meets a person with a red nose. What is the probability that he will be a liar?'
B 'Pingping goes to a village to ask for directions. In this place 10 villagers out of every 100 will lie. Of the 10 people who lie, 8 have a red nose. Of the remaining 90 people who don't lie, 9 have a red nose. Imagine that Pingping meets a group of people with red noses. How many of these people will lie? ____ out of _____?'

Zhu and Gigerenzer contrast two ways of solving this problem. One is to apply Bayes' well-known formula to these probabilities. The formula for a binary hypothesis (*H* = liar, *not-H* = not a liar: *D* = data) is:

$$p(H|D) = \frac{p(H)p(D|H)}{p(H)p(D|H) + p(not\text{-}H)p(D|not\text{-}H)}$$

Applied to the story, this equation would be:

$$p(H|D) = \frac{.10 \times .80}{.10 \times .80 + .90 \times .10}$$

$$= .47 \text{ (to 2 decimal places)}$$

The second way is to proceed through the story working out the frequencies: among the 10 liars 8 will have a red nose: among the 90 non-liars 9 will have a red nose: thus there will be 17 people with red noses, and 8 out of the 17, which is just under half or $p = 0.47$, will be liars.

The researchers report that when the quantitative information was given as probabilities, none of the children solved any of the problems or even attempted to solve them in a way that made any sense. 'The children' they remark 'seemed to have no clue how to solve the problem when the information was presented as probabilities'. In clear contrast, the 9-year-olds succeeded a modest 14% of the time when the information came in the form of frequencies, the 10-year-olds 42% of the time and the 11-year-olds 47% of the time. The difference between the two conditions is quite remarkable, especially given that the probabilities in the impossible condition were presented as percentages, since percentages are highly similar to frequencies. It should have been quite easy to transform percentages into frequencies by treating, for example, the statement 'If a person lies, the probability that he has a red nose is 80%' as 'Eight out of ten liars have red noses', but apparently the children did not make this move. Yet, the relatively high scores in the frequencies condition, especially among the oldest children, demonstrate that many of them were able to devise an effective and appropriate solution that involved a sequence of quite complex logical steps.

One possible explanation for the children's successes in the frequencies condition is that the information that they were given could have led them to concentrate on the absolute numbers of the different elements in the sample space. The data they are given leads them directly to the calculation that there are 8 liars and 9 non-liars with red noses. In our view, this would fit well with the argument that children solve proportional problems well when they have a clear idea of the quantities of the different elements in the sample space. Knowing about these should help children who are thinking in terms of ratios.

Whatever the reason for the stark difference between the two conditions, it is valuable to note that Gigerenzer and his colleagues (Gigerenzer, 2002; Hoffrage and Gigerenzer, 1998; Hoffrage *et al*., 2002; Hoffrage *et al*., 2000) found a similar difference in adults to whom they gave this sort of conditional probability problem. This was true even when the problems were about illness and the researchers gave them to medical professionals.

Summary of the research on conditional probabilities

1. Conditional probability problems are hard for children and for adults when the information provided is in a proportional form. However, the same problems become easier when the actual quantities in the sample space are given. In our view these results fit well with the hypothesis that children are at their best in solving probability problems when the information that they deal with is about the absolute numbers in the sample space, and they can make these into ratios.

2. The argument that we have developed about children's ability to understand ratios and to use them to solve probability problems is relevant to the way probability is taught at school. The idea should be pursued in intervention experiments.

5. Correlations

Between certainty and uncertainty

Between the complete certainty of determined events and the complete uncertainty of totally random events lies a world of imperfect, but nonetheless important associations. This is the world of associations between variables. Correlations are measures of the strength and the direction of the association between two variables. A correlation coefficient of 1 tells us that two variables are perfectly and positively related and a coefficient of <minus>-1 shows that they are perfectly but negatively related. Neither of these correlations leaves any room for uncertainty, but these perfect correlations are extremely rare. Correlations are greater or smaller than 0 and fall somewhere between 0 and 1 or between 0 and <minus>-1. They show that there is an association between the two variables but also indicate that we cannot be certain how the association will affect individual cases. We know, for example, that there is a relationship between how much people eat and whether their mass goes up or down, but we also know that the association between these two variables is not a perfect one, since the strength of the effect varies a great deal between people. The association, though less than perfect, allows doctors and dieticians to give good and worthwhile advice to people in danger of obesity, for example, but it is not strong enough for them to make precise predictions about what will happen to individuals as a result of changing their diet.

Many situations that confront us, both in science and in everyday life, involve at the same time associations between variables and some uncertainty about the effects of the association. Correlational reasoning is about the presence, nature and strength of a mutual relationship between two variables (Adi *et al.*, 1978). This reasoning requires the recognition that relationships between variables are not absolute but exist in degrees (Ross and Cousins, 1993), and thus involve probabilistic reasoning.

Risk is an uncertainty that ‘can be expressed as a number such as a probability or frequency on the basis of empirical data’ (Gigerenzer, 2002, p. 26). In everyday life the word ‘risk’ is associated with negative outcomes, but this is not the use made of the word in the medical or psychological literature. To use a common example, in everyday life one would speak about the risk of having cancer (i.e. the probability of having cancer) if a test is positive but one would not speak of the risk of not having cancer (i.e. the probability of not having cancer) if the test is positive. However, the probability of not having cancer if the result of the test is positive is still important and must be considered when decisions are made about the subsequent course of action. According to Gigerenzer’s definition, each of these probabilities defines the risk of that event taking place.

The word risk carries another connotation, alongside the probability of an event, which is related to the seriousness of an outcome. Pascal explored this meaning of risk in the argument that is today known as Pascal’s wager (Mlodinow, 2009). Pascal weighted the probability that God existed, if one did not know anything to prove it one way or the other, against the severity of the risks one could run by following or not the laws of God. Pascal’s wager is presented here in a simpler form, focusing on the meaning of risk in terms of severity of the outcome. The risk of following the laws of God is that one might miss some pleasures in a life of limited in duration. The risk of not following the laws of God is losing eternal life and happiness. So Pascal concluded that every reasonable person should obey the laws of God because the risk associated with not obeying them, if God exists, is clearly more serious than the risk of obeying

them, if God does not exist. Although this connotation of the work risk is important, the focus in this paper is on the first one, the probability of an event taking place.

Correlations help people define the probability of a particular event taking place when something else is known. We can take, as an example, a committee carrying out an inquiry into the deaths of children undergoing a certain type of surgery in a particular hospital. The committee must consider the evidence using correlational reasoning. The question the committee needs to answer is whether children operated on in this hospital are more likely to die than those who received the same surgical intervention in other hospitals. In other words, is there an association between receiving the treatment in this hospital and death? The relationship is unlikely to be absolute: not all children operated in the hospital will have died and not all operated elsewhere will have survived. But the question of whether the chances of the children dying are increased by receiving surgery in this particular hospital can still be asked and answered by looking at the strength of the association.

The value and importance of these imperfect associations is now widely recognised. Even the most impressive discoveries in the history of science might have been dismissed if one were to expect a perfect association. When Florey and his colleagues ran the first experimental study on the effects of penicillin with mice (Lax, 2004), the outcome could be interpreted in deterministic terms: they infected 8 mice with the smallest dosage of virulent streptococci known to kill a mouse of average mass and then gave four of them penicillin. The four mice that did not receive penicillin died within a day. Of those that did receive the penicillin, two had received a single shot, and one died after two days and the other after six days. Of the two that had received five shots over a period of time, one died after 13 days and the other lived on, presumably to a ripe old mouse's age. However, the trials with humans were by no means as clear-cut: two of the first six patients treated with penicillin died, which suggested the possibility that penicillin may not be as great a success as the mice experiment had suggested. The question was whether this recovery rate was definitely better than a no-treatment condition, and Florey certainly wanted to seek more evidence before making these experiments known to the world.
Thus correlational reasoning is an essential element in scientific reasoning and scientific literacy (Gigerenzer *et al*., 1989; Robinson, 1968; Ross and Cousins, 1993), and a means of controlling the present and predicting the future in order to maximise the desired outcomes in one's personal life (Alloy and Tabachnik, 1984).

The cognitive demands involved in understanding correlational reasoning

The cognitive demands in understanding correlational reasoning are various, and we focus here briefly on three. The first one is understanding randomness. If there is no relationship between two events, A and B, it is still possible that they might occur together by chance. The aim of analysing the correlation between two events is to establish whether they co-occur more often than one would expect by chance. Understanding randomness is therefore part of understanding correlations.

Correlational reasoning also involves understanding sample space. In order to examine whether two events are associated, we need to establish not only whether they co-occur but also what all the possible cases are: Did A happen? Yes or no. Did B happen? Yes or no. The sample space here is four possible categories.

1. Yes–Yes
2. Yes–No
3. No–Yes
4. No–No

One could be tempted to think that only the Yes–Yes cases are relevant to the question of a correlation between the two events, but the probability of the events occurring together must be understood in the context of the events not occurring together as well.

If we think of the infected mice that did or did not receive penicillin, we have a slightly more complicated sample space. Some mice did not receive penicillin and did not survive for one day, so this is a No–No case (i.e. no penicillin, no survival). Some mice received a single shot of penicillin: one survived two and the other six days. If we simplify the survival criterion to, for example, surviving one week, these mice would also be Yes–No cases. Some mice received five shots of penicillin and survived for longer than one week: they would be examples of Yes–Yes cases. There were no mice that exemplified the No–Yes case (i.e. infected mice that did not receive penicillin and survived for at least one week).

This example hints at how sample space can be much more complicated than cases in four categories, because the cases may vary in more subtle ways than Yes or No. One could, for example, characterise the administration of penicillin by the number of shots the mice received and the survival of the mice by the number of the days they survived. This would create a much more complicated sample space, which cannot be analysed as easily. However, most of the research on children's understanding of the association between variables has focused on the simplest sample spaces described by Yes or No on the two variables.

Once we have established the sample space, we need to move on to the quantification of probabilities in a proportional manner. Is the frequency of cases that support the existence of an association (the Yes–Yes and the No–No cases) proportionally really larger than the frequency of the cases that do not support the association, so that one can assume that this frequency departs from what one would expect by chance? If this is the case, we conclude that there is an association between the two events.
In summary, understanding the association between two variables makes at least three demands on children's reasoning: they need to understand randomness and the sample space, they must be able to recognise which cases support and which cases go against the idea of an association between the variables, and they need to be able to quantify the positive in comparison to the negative cases in order to assess whether the positive cases are frequent enough to suggest that the co-occurrence observed is not due to chance.

Past research on how people understand correlations

Studies of people's understanding of correlations can be best analysed if we separate them into groups, defined by the questions that they address. This is not to say that the questions are independent of each other; it is just a methodological step to help us identify the different pieces of the puzzle that constitute correlational reasoning. The studies are sorted in this report into five types related to the questions they address.

1. How do people react to contradictions of an expected relationship between two events?
2. What sort of information do people seek when trying to find out whether two events are correlated?
3. How does the presentation of information relate to our understanding of correlations?
4. How do children and adolescents quantify the information that they are presented with and what inferences do they draw from information?
5. How can we help students to understand correlations better?

Reaction to information that contradicts an expected relationship between variables

Inhelder and Piaget (1958) carried out a variety of studies in which they analysed how children and adolescents react to information that confirms or contradicts the existence of an expected relationship between variables. In one well-known study, they asked children to attempt to explain why things float or sink in water. This situation is a deterministic one, and not a matter of probabilities, and the researchers' interest was in the way children reacted to contradictions of their predictions. The problem is rather appropriate for examining reactions to contradictions because many people, including many adults, start out with the notion that heavy things sink and light things float, rather than with the idea of density, which involves a relation between mass and volume. The question is then how participants will react to the contradiction of their predictions. If children and adolescents cannot discard the hypothesis of a relationship between mass and sinking (i.e., if they cannot reject the hypothesis that heavy things sink and light things float) in a deterministic situation, they might find it even more difficult to interpret relationships that are probabilistic rather than deterministic. We focus here on the relevant aspects of this study, not on the details of whether and how the participants reached an understanding of density.
In the Inhelder and Piaget study, at the start of the session, the children are asked to classify the objects in two categories, those that will float when placed in a basin full of water and those that will sink when placed in the basin. If a child forms these two categories and provides a consistent explanation – for example, these float because they are light or small and those sink because they are heavy or large – the experimenter proceeds to ask for specific predictions for each object and then notes the child's reactions to contradictions of these predictions. The experimenter chooses, for example, a large piece of wood, which is both large and heavy, and asks the child to make a prediction. To be consistent, the child should predict that it would sink. When the wood is put into the basin, it floats. Inhelder and Piaget noted three different types of reactions to this contradiction.

1. Some children would ignore the contradiction, and continue to assert that heavy things sink and light things float and indeed attempt to make the piece of wood conform to their prediction by pushing it down to the bottom of the basin.

2. Other children would modify their hypothesis, forming classes of objects that can float despite belonging to a class which is predicted to sink (e.g. heavy objects sink but wood normally floats because it has air inside, it is not very compact).

3. Other children would note the contradiction and would no longer accept the simple association between mass and sinking (some of these actually give up seeking the solution, whereas others seem to go on to think of a relationship between mass and volume, constructing an understanding of density).

The relevance of this study to correlational reasoning may not be immediately apparent but we hypothesise that it is not possible to think about correlations without understanding how expected relationships might be disconfirmed by evidence.

Inhelder and Piaget's (1958) work on propositional reasoning, exemplified in the study about the law of floating bodies and the elimination of contradictions, inspired a large number of subsequent studies on how children and adolescents interpret statements about causal relationships and how they interpret contradictions. It should be noted that studies on contradiction do not imply that children have no understanding of causality. For example, children know that if water is spilled, the floor gets wet, and if something is cut, it is no longer in one piece (Bullock and Gelman, 1979; Schultz, 1982, das Gupta and Bryant, 1989; Sobel and Kirkham, 2007). Inhelder and Piaget's studies were about how children re-examine their thinking about relationships if their thinking is contradicted by observations. As far as we know, other researchers have not disputed Inhelder and Piaget's central claim that children's ability to see the relevance of disconfirmation of a prediction about a relationship between two events improves over time and that this ability is not observed among children in the early years of school. This important result has consequences for understanding correlations: if children cannot discard their explanations for events when these are contradicted in a deterministic situation, it will be difficult for them to do so in a probabilistic situation, in which both cases that confirm and cases that lead to disconfirmation of the prediction might be observed.

Subsequent research has not analysed reactions to contradictions in such detail but has explored the judgements that people make about correlations depending on whether they hold beliefs about the association between the variables. There are several studies on the recognition of covariation between variables when participants have a certain bias (e.g. Alloy and Tabachnik, 1984; Jennings, Amabile, and Ross, 1982; Scholz, 1991) but the best-controlled study was by Batanero *et al*., (1996). Batanero and colleagues presented to a large sample of final year secondary school students (age 17 to 18 years) tables (2 <mult> x 2, 2 <mult> x 3 or 3 <mult>x 3) that contained frequencies showing the co-occurrence of certain characteristics. For some of these, they expected the participants to have previous beliefs – for example, an association is expected between smoking and having a bronchial disease and between number of hours studied and results in an exam. For other characteristics, the students were not believed to have expectations: for example, leading a sedentary life and having a skin allergy. The students were asked to interpret the tables and answer whether there was an association between the variables. The level of difficulty of the problems, as defined by the size of the contingency tables and the direction of the association (direct relations are more easily recognised than inverse relations), was controlled across conditions of expectation. Batanero and colleagues found a strong association between the prior beliefs of the students and their interpretations of the contingency table. Even students who correctly analysed the proportions of confirming cases and the proportions of disconfirming cases often drew the wrong conclusion, either supporting the association when there was none according to the table or failing to detect an association when it should have been detected. Because these were secondary school students, these results suggest that the interpretation of information about correlations is influenced by reactions to contradictions, even though the same participants

might have reacted differently if the contradiction had been to a prediction in a deterministic situation.

Seeking information about relationships between events

Among the numerous studies that were inspired by Inhelder and Piaget's work on propositional logic, one set of studies focused specifically on the analysis of how people seek information in order to test whether an association between two events affirmed in a proposition is true or false. Wason (1968) designed a task in which participants were asked to test whether there was an association between what was written on one face and on the other face of a set of four cards. The association was presented to the participants as a rule: 'If there is a vowel on one side of the card, there is an even number on the other.' The participants were asked to select only the necessary and sufficient pieces of evidence to test whether the rule is true. The cards that are on the table show a vowel, a consonant, an even number and an odd number. The necessary and sufficient information in this deterministic situation is to select the card with the vowel and the one with the odd number, because either could disconfirm the rule; the card with a vowel can also provide confirmatory evidence. The cards displaying a consonant and an even number are seen as irrelevant to the rule, because the rule does not state that there is a mutual association between even numbers and vowels.

The commonest behaviour by children and adults in this task is to choose to verify what is on the other side of the card with the vowel and of the card with the even number. This choice is considered an error in testing the correctness of the rule because the card with the even number could not lead to disproving the rule. The participants' behaviour in this situation has been interpreted as revealing what has come to be known as a confirmation bias, i.e. a search only for information that would lead to confirming the rule without a realisation that other information could result in disconfirming the rule.

Subsequent research (e.g. Cheng and Holyoak, 1985; Cheng *et al.*, 1986; Girotto, Light, and Colbourn, 1988) sought to provide an alternative interpretation to the behaviour of children as well as of adults in this task. Their behaviour was considered not to be adequately described by a logical analysis but rather by pragmatic schemas, which determined the relevant cases to be analysed. If the 'if–then' statement was interpreted as a permission or a prohibition, rather different behaviour in the testing of rules was observed.

These studies suggest that children and adults evaluate the relationship between events differently depending on what they expect the nature of this relationship to be. Therefore, in studies about children's understanding of correlations, which are mutual relationships between events, one must ascertain whether they understand what a mutual relationship means when they test its existence. If their behaviour seems to indicate, for example, a confirmation bias, as in the Wason four-card problem, we need to consider what consequences this bias has for the understanding of correlational reasoning.

'Confirmation bias' is a term used to refer to seeking evidence that can only support the existence of a presumed association between two events. This bias does not have to be intentional or explicit: Nickerson (1998) defines confirmation bias as 'unwitting selectivity in the acquisition and use of evidence' (p. 175). Evans (1989) consider this as 'perhaps the best known and most widely accepted notion of inferential error to come out of the literature on human reasoning' (p. 41) and Dawes (2001) suggests that professionals may be prey to this

bias as a consequence of their professional experience: for example, if a clinical psychologist asserts that child sex abusers do not recover from this condition without professional help, this assertion is often based only on the cases that the professional has seen, namely those who seek professional help. Disconfirming cases, i.e. those who recover from their condition without professional help, are usually not part of a psychologist's experience.

Loren Chapman and Jean Chapman analysed confirmation bias in a series of studies with psychologists and undergraduate psychology students who were given information about patients and also about their performance in psycho-diagnostic, projective tests. In one study (Chapman and Chapman, 1967), for example, the participants were presented with drawings of human figures supposedly produced by patients who had one of six symptoms (e.g. he is suspicious of other people, he is worried about how manly he is). The pairings of the drawings with the symptoms had been carried out randomly. However, the participants supposedly discovered relationships between characteristics and symptoms as a consequence of remembering only confirmatory cases: for example, 80% of the participants 'discovered' that a figure drawn as muscular, with broad shoulders, indicated that the patient was worried about his manliness. Chapman and Chapman referred to this as illusory correlation, which they describe as the erroneous reporting of co-occurrence of symptoms and signs in a diagnostic test (see also Chapman, 1967; Chapman and Chapman, 1975).

In summary, research with adults and children has suggested that they are influenced by the nature of the relationship that they expect to exist between two events in the way they search for information to test whether the relationship exists. Some researchers have described a confirmation bias or illusory correlation in a number of situations and by a variety of participants – but note that these are simply terms and do not constitute an explanation for why information is selected in a particular way. Chapman (1967) attempted to explain this phenomenon as a consequence of stronger memory for associations between phenomena that were previously associated in one's experience. However, the information does not have to be committed to memory for this bias to be observed. In subsequent descriptions of how children and adolescents deal with information about the relationship between two events, we will consider the possibility that the confirmation bias stems from cognitive demands made by tasks and the difficulties that we have in dealing with such tasks.

Presentation of information and understanding correlations

In order to assess whether two events are correlated, people must be given information. It could be argued that the best way to analyse correlational reasoning is to provide information to participants about individual cases because organising the information can be seen as part of understanding how to assess whether two events are related. Mlodinow (2009) actually suggests that historically the analysis of probabilities only became possible when people developed better means of recording the occurrence of events.

Inhelder and Piaget (1958) briefly mention, in their study of adolescents' correlational reasoning, that the participants performed better when the classes of events were presented to them in 2 <mult>x 2 tables, which organised the information according to the sample space – Yes–Yes, Yes–No, No–Yes, and No–No.

Other researchers have shown that children may have difficulties in sorting out the information in order to construct the relevant classifications for a table (Adi *et al*., 1978) and that providing

children with information already organised in tables improves their performance in tasks in which they are asked to assess whether there is an association between two events (Carvalho, 2008). Ross and Cousins (1993) showed that students can be taught how to organise information about individual cases in 2 <mult sign> x 2 tables, which can then be scrutinised in order to assess whether the events are associated. They also showed that students can be taught to organise information even in more complex, multivariate situations, in which a relationship between variables only exists under one condition but not under another (e.g. the relationship between a treatment and recovery may be conditional on the amount of medication used). Ross and Cousins found that the ability to organise the information in tables can be considered part of the skills necessary for correlational reasoning because some students cannot even start to organise the information. However, this ability may improve without a similar improvement in the ability to make correlational inferences from the information. Thus organising information can be seen as an important step in assessing correlations, but is distinct from the process of making inferences about whether a correlation does or does not exist. Tables still have to be analysed in order to assess whether there is a correlation between the variables.

Although tables are often used in correlational reasoning studies in which children and adults are asked to assess whether there is a correlation between events, information about correlations is often presented in the media and in scientific papers, not in tables but either in conditional probabilities or in ratios. Probabilities can be stated as percentages and proportions or as ratios, referred to by some researchers as frequencies. Scientific and media reports seem to resort to proportions or percentages because these figures are easier to compare than frequencies: for example, if we are told that 62 people in a sample of 243 from city A had a particular illness and that 93 people in a sample of 329 from city B had the same illness, it is difficult to know whether the illness was relatively more frequent in city A than in B. If we were told, in contrast, that approximately 25% of the people in each sample had the illness, we would quite easily conclude that the incidence was very similar in the two cities. The ease with which we compare figures in this example does not extend to correlational reasoning, in which the percentages or frequencies are related to conditional probabilities. When information is presented in frequencies or ratios rather than percentages or proportions, both children and adults seem to find the information easier to interpret.

Correlations can be presented in 2 <mult sign> x 2 tables, when the events are discrete (of the Yes–No type) and, when the variable are continuous, they can be presented either in tables or in graphs or actually in both formats at the same time. Carvalho (2008) analysed secondary school students' inferences about specific data points or trends in co-variation situations and found that they did not appear to use information from graphs by looking at the spatial characteristics of the graphs alone: when they explained their answers, they did not refer to slopes, for example, but to values that they read from the graphical representation. Their performance in problems presented in tables or both graphs and tables did not differ significantly, probably because they relied on numerical information. When the graphs represented a negative correlation, secondary school students found it rather difficult to draw the appropriate inferences from the graphs, although they could note the negative relationship between the variables when the information was presented in tables.

The quantification of probabilities to determine whether there is a mutual relationship between two events

The initial work on how children use information to decide whether two events are related was carried out by Inhelder and Piaget (1958). They asked the adolescents aged 12 to 14 years who participated in their study to ascertain whether there was an association between hair colour (blond hair versus brown hair) and eye colour (blue eyes versus brown eyes) in a set of cards. The researchers made it clear to the participants that the question referred simply to the set of cards presented to them, in which faces with these attributes were drawn, and not to their experience outside the cases they were considering.

The researchers were initially interested in examining the responses from the standpoint of propositional logic, and not in the quantification of the relationship. The sample space (or possible cases) was thus definable as Blond–Blue eyes (p and q), Blond–Brown eyes (p and not q), Brown hair–Blue eyes (not p and q), and Brown hair–Brown eyes (not p and not q). The researchers asked the participants whether there was a relationship between hair colour and eye colour in this set of cards and later to subtract or add cards that would make this relationship stronger or weaker.

Inhelder and Piaget noted that participants had a problem right from the start with establishing how the four classes that define the possible cases (or the sample space) relate to the question of whether there is a relationship between eye and hair colour. They tended to establish one class – for example, the class Blond–Blue eyes (p and q) – and base their answer on this class only, thereby thinking only of the probability of having blond hair and blue eyes in that sample. When the researchers made the distribution of cards such that the relationship was actually perfect – 6 cards with blond –blue eyes (p and q), 6 with brown hair and brown eyes (not p and not q) and zero cards in the other categories – some adolescents dissociated the two categories from each other in their answers. The blond hair–blue eyes cases indicate that you have more chance to have blue eyes if you are blond; when asked about the brown hair–brown eyes cases, they thought this was not relevant, these cases only indicated that you are more likely to have brown hair if you have brown eyes.

Other adolescents realised that both of these classes are confirming cases and the remaining classes are disconfirming cases, but did not combine the confirming cases in order to compare them to the disconfirming cases as a group. When they were asked to compare two sets of cards and say in which set they were more likely to find cases that ‘follow the rule’ of a relationship between hair and eye colour, they did not use all four classes in their answers. For example, when comparing two distributions with the same number of cards with blond hair–blue eyes, the same number of ‘errors’ according to the rule, and different numbers of cards with brown hair–brown eyes, they thought that the relationship between hair and eye colour in the two sets is the same: there were five chances of being right in both sets (ignoring that the brown hair–brown eyes increased the chances of being right in one of the sets).

Inhelder and Piaget suggested that, with further comparisons, as the experiment proceeded, some adolescents were able to reach the understanding that they needed to relate the sum of the sets of confirming cases to the sum of the sets of disconfirming cases. They also suggested that only exceptionally the adolescents anticipated the need to combine the two types of cases in their analyses. Thus, although some adolescents were able to reach an understanding of how cases confirming and disconfirming the relationship could be taken into account quantitatively,

they did not come across many that demonstrated this understanding from the outset of the experiment.

It should be pointed out that in this study, as in the study on quantification of probabilities, Inhelder and Piaget's clinical method may lead to a more positive assessment of adolescents' understanding of probabilities that other methods, in which the participants are asked questions but not presented with conflicting approaches to the problem or asked to increase or decrease the relationship between the variables by manipulating the number of cards in the different cases.

The study by Inhelder and Piaget inspired some further research, both with adolescents and with adults. Much of the work with adolescents consisted in developing measures to assess correlational reasoning (e.g. Tobin and Capie, 1981), using such measures to predict students' success in science courses or assessing whether science courses had an effect on correlational reasoning (Lawson, Adi, and Karplus, 1979).

Studies with adults focused on whether adults who did not take many mathematics courses showed competence in correlational reasoning in their domain of work. Smedslund (1963), for example, interviewed nurses and student nurses (N = 96; in Denver and in Oslo) using a task which was very similar to the one used by Inhelder and Piaget, but the relationship to be analysed was relevant to their work: it concerned the correlation between a symptom and an illness. They were instructed to concentrate only on the information in the cards, which were meant to be about patients, numbered from 1 to 100 in the order in which they were admitted to the hospital. The cards contained four letters, representing different specific symptoms, and four other letters, representing specific diagnoses made by the hospital. The nurses were asked to focus on each of the associations, one at a time.

Smedslund expected that participants that understood correlational reasoning would use a selective strategy, organising the cards into the four categories relating to presence or not of the symptom and presence or not of the diagnosis; would count or estimate the frequencies in each category; would attend to all four categories; and would compare the frequencies of the sum of confirming cases with the sum of disconfirming cases. Some of the participants were also shown the frequencies for the different combinations of presence or absence of the symptom and the diagnosis and asked to estimate the strength of the relationships between symptoms and illnesses.

Smedslund found some of the same behaviours described by Inhelder and Piaget in these participants: some nurses seemed to think that a relationship may exist when the symptom and the diagnosis co-occur and at the same time may not exist when neither is present in the cards; they did not see both cases as examples supporting the mutual relationship between the symptom and the diagnosis. If a symptom, such as a headache, appeared in many illnesses, it was considered an important factor for a diagnosis of a particular illness even if it appeared on the card only sometimes. Smedsland reported that not a single participant gave an indication of having understood that the degree of the relationship between symptom and illness depended on the ratio of confirming to disconfirming cases. This study, as some of those referred to in the section on quantification of probabilities, suggests that the clinical method used by Inhelder and Piaget in their interviews may have led to a positive view of what adolescents achieve, which is not replicated in other studies in which less interactive methods are used.

The concern with lack of evidence of correlational reasoning in secondary school and university students documented in different studies (e.g. Karplus, Adi, and Lawson, 1980; Lawson, 1982) inspired efforts to develop ways of promoting this reasoning, exemplified by studies summarised in the subsequent section.

Improving students' understanding of correlations

Researchers interested in promoting the development of correlational reasoning, such as Nous and Raven (1973) and Ross and Cousins (1993) seemed to approach correlational reasoning as involving a set of interrelated skills, such as organising data in tables or graphs, articulating predictions (e.g. in a graph, if the value of variable A increases, does the value of variable B increase, decrease or stay the same?), locating data, synthesising data, and drawing conclusions. The results suggested, as briefly mentioned above, that it is possible to improve the skills of organising and synthesising data without improving the inferences regarding the correlations. Students in the taught groups were able to organise and synthesise the data better than those in the untaught groups; however, they continued to answer the correlation questions, not by using the information, but on the basis of their preconceived ideas. For example, one of the questions was whether taller people were faster swimmers. Students in the taught groups could say either 'yes' or 'no', and did not reach a similar conclusion, even though they had organised and synthesised the data appropriately; when drawing their conclusions, their arguments were not based on the data. The actual intervention in this study was carried out by classroom teachers, and Ross and Cousins (1993) report a strong relation between the commitment of the different teachers and the effect of the intervention on the students' achievement. They also found that their programme led to similar results with younger, grade 7 students (about 13 years of age), and older, grade 10 students (about 16 years of age). Other approaches to teaching correlational reasoning also used lessons on correlations, but these were combined with teaching other formal operations schemes, such as control of variables and proportional reasoning (e.g. Lawson and Snitgen, 1982). Although these are relevant and important studies, it is difficult to relate the improvements in correlational reasoning to a specific aspect of the teaching programme.

As far as we know, only one recent teaching study by Vass, Schiller and Nappi (2000) conceived of the process for promoting correlational differently from the earlier studies. The early studies approached the teaching of correlational reasoning as involving a set of skills, as described in the Ross and Cousins study. Vass, Schiller and Nappi's concentrated on the conceptual basis on which correlational reasoning rests. Inhelder and Piaget (1958) and Karplus, Adi and Lawson (1980) advanced the hypothesis that understanding correlations depends on two cognitive schemes: understanding probabilities and understanding proportions. Vass and colleagues taught one group of teacher-education students in three lessons about proportionality and probabilities only and a second group, also in three lessons, about proportionality, probabilities and correlation. The first two lessons were the same for both groups but the third one differed. The members of the first group were given a review of the concepts of proportionality and probabilities correlations in the third lesson. The second group was taught about correlations in this lesson. Vass *et al.* reasoned that, if proportionality and probabilities are building blocks for understanding correlations, the group taught only about these two concepts should make progress in understanding correlations, but perhaps not as much as the group that was also taught about correlations.

Both intervention groups made more progress from pre- to post-test than the control groups. The means attained in the correlational reasoning measure by the two groups were almost identical at post-test, and for both groups significantly better than their pre-test performance. Vass *et al.* concluded that teaching them about the building blocks gave them the necessary schemes to reason about correlations even without specific teaching about correlations. This is an impressive demonstration of how helping students to meet the separate demands of a complex concept can promote significant advances in the students' understanding of that concept.

Summary and conclusions

We argued at the start of this section that correlational reasoning involves the co-ordination of three schemes of reasoning: understanding randomness, the sample space and quantification of probabilities. Correlational reasoning goes beyond each of these on their own and provides the basis for much of modern science and for understanding a large variety of situations in everyday life. Many of the relationships between events and between variables are not deterministic, but are probabilistic in nature. In order to assess whether there is a mutual relationship between events, we must test whether the association that we note between their frequencies is one that departs from what could be expected by chance. We must also understand the sample space that allows for scrutinising the relationship: in a typical discrete variable situation, we can characterise this space as the combination of Yes and No for each of the events. For example, a plant was treated with a pesticide (Yes–No) and it is no longer infested (Yes–No), gives a table with four cells: Yes–Yes, Yes–No, No–Yes, and No–No. In order to draw inferences from the frequencies in this table, we must understand the relevance of the different cells to a mutual relationship between the variables and draw conclusions accordingly. Two of the cells represent confirmations of the mutual relationship (Yes–Yes and No–No) and the other two represent disconfirmations. We then need to know how to deal with these cells quantitatively.

Reasoning about a contradiction to a hypothesis is not simple in deterministic situations; it is also not simple in correlational situations. Some of the studies we reviewed, including for example the Smedslund study with adults, show that contradictions might not be noted by them when considering correlational evidence: they might focus only on the cases that seem relevant to them (the Yes–Yes cell in the 2 <mult> x 2 table) or they might think that if two events co-occur sometimes, then they must be associated in some way. Organising the information is important for scrutinising a relationship between events but even when this is not necessary, because the cases are already organised and quantified in the categories, drawing inferences is still a difficult matter.

The quantification of the confirming and disconfirming cases in order to test for a mutual relationship between two variables was rarely observed by Inhelder and Piaget at the outset of their experiment. Even their clinical method, which tended to guide participants to think about problems from different perspectives, did not suggest that correlational reasoning is easily attained. Subsequent research indicated that many secondary school and college students might not demonstrate high levels of correlational reasoning. These results motivated the search for methods of improving correlational reasoning, which was seen as particularly important in biological and medical sciences, and which indeed predicted students' success in courses in these domains.

Teaching studies that aimed to improve secondary school and college students' correlational reasoning were sometimes conceived in terms of the skills necessary for correlational reasoning and sometimes were part of a larger programme designed to improve formal reasoning in general. We found a single study that was conceived differently: a control group was compared to two taught groups, one that had received instruction only on the schemes necessary for correlational reasoning (probabilities and proportionality) and one that had received instruction on these two schemes plus instruction on correlational reasoning. The striking similarity in results in the two taught groups provides strong support for the notion that, if people master the cognitive demands of correlational reasoning, they can use these resources in order to understand correlational problems. It is, however, too soon to reach a firm conclusion, on the basis of a single study. A combination of longitudinal, predictive studies showing that these particular demands predict learning of probabilities (even after controlling for other intellectual measures) and further teaching studies is required to support our hypothesis that correlational reasoning is the reward that one can reap in the scientific domain from understanding randomness, sample space and proportional quantification in the mathematics classroom.

6. General summary

In the introductory section we also argued that, although much of the research on children's understanding of probability is based on good ideas and on ingenious tasks, the design of much of this research is very limited. The concentration on cross-sectional studies and the near-complete absence of longitudinal research designed to test hypotheses about the connections between the strength of various abilities and the progress that children make in reasoning about probability is one serious problem. Another is the scarcity of intervention projects that are designed to test causal hypotheses about the factors that affect children's learning about probability.

In the section on randomness, we reported evidence that children at first have difficulties in distinguishing random from determined events, but do acquire relevant and reasonable ideas about randomness, particularly in the context of fairness, by the age of about 10 years. The relations between fairness and randomness has been under-researched, except in work on computer microworlds which suggests that everyday experiences like shuffling cards and tossing coins help children learn about the importance and the consequences of randomness. This idea is relevant to the teaching of probability at school, and should be explored in intervention studies. We also discussed work on infants under 1 year in age which has led to claims that even these young children can distinguish events that happen at random from events that are determined because the actors concerned can make informed choices. This research is ingenious, but some serious questions about the design and procedures of these experiments will have to be answered before we can accept the striking conclusions that the research has led to.

The main point that we made about the sample space was its central importance in understanding probability and in solving probability problems. Yet, research on children has produced very little direct information on how children learn about the importance of the sample space or about how to analyse it. Much of the information that we have is negative, since it comes from mistakes that children make in reasoning about probability, which they wouldn't have made if they had had a thorough grasp of the sample space. On the positive side, there has also been research, starting with studies described in Piaget and Inhelder's book on

chance, on children's combinatorial reasoning. This is relevant to the analysis of the sample space because in many probability problems one has to aggregate the elements of the sample space in combinations and categories. The innovation of children working with computer microworlds on creating sample spaces is also impressive: microworlds appear to be an excellent way of dealing with the large amount of data that is often part of discussions about sample spaces.

Much of the evidence on children's combinatorial reasoning suggests that it is often difficult for children to make exhaustive lists of all possible combinations and compounds within a sample space. However, research on Cartesian product problems establishes that quite young children can often manage the exhaustive reasoning that is needed for these problems if they can model the space with concrete material. This would be valuable information for anyone devising an intervention study of how to teach children about the sample space. Intervention studies are badly needed, partly for educational reasons and partly to test hypotheses, like Piaget and Inhelder's, on the importance of combinatorial reasoning for learning about probability.
The section on the quantification of probability began with a discussion of the relatively few probability problems that do not depend on proportional reasoning for a solution because they can be solved on the basis of simple more/less relations. The recent experimental work on infants' knowledge about probability is based on such problems, and the generally positive results of these experiments raises the question whether babies would also manage well when given genuinely proportional problems. The evidence on schoolchildren's attempts to calculate the probabilities of simple events, and to compare probabilities across different sample spaces, suggests that the need to reason proportionally in these tasks is a real obstacle for them. However, some of the reports suggest that when children of 9 to 12 years do succeed in these problems they usually do so on the basis of working out the ratios between different numbers in the sample space rather than by working with fractions. The possibility that children work with ratios better than with fractions to solve probability problems may also be the underlying reason why they are much more likely to solve conditional probability problems when they are given absolute rather than proportional values for the elements in the sample space.

In the section on correlations, we dealt with children's reasoning about situations in which two variables are significantly, but imperfectly, associated. Correlations are therefore about the probability of a relationship, and correlational reasoning diverges from reasoning about deterministic connections in many ways. Correlational reasoning involves dealing both with confirming and with disconfirming evidence in order to quantify the strength of the association, and the degree of uncertainty that it entails. Research on students' reactions to correlational problems shows that these are generally difficult even for people in their final years at school. However, there is now some evidence that children's correlational reasoning improves as a result of teaching about the underlying concepts of proportion and probability.

In the report, we make two main recommendations about research on children and probability. The first is to take advantage of research designs that have been successful in research on other aspects of children's intellectual development. In particular, we recommend the combined use of intervention and longitudinal methods to study the links between the four aspects of probability that we have discussed in this report.

The second recommendation is that researchers on children's understanding of probability should pay much more attention than they do at the moment to the great amount of related data on other aspects of cognitive development. Probability, as we have seen, makes a number of

different cognitive demands and most of these demands are shared with other aspects of cognitive development about which we know a great deal. Probability is an intensive quantity, but so are density and temperature. Analyses of the sample space require combinatorial reasoning: so do many branches of scientific thinking. We think that many people doing research on probability have paid very little attention to research on these related topics, and have missed out on potentially valuable information.

Nevertheless, research on children's understanding of probability is a thriving concern. It continues to produce interesting ideas and striking empirical results. It deserves a great deal of attention and encouragement.

.....

References

Abrahamson, D. (2006) The shape of things to come: The computational pictograph as a bridge from combinatorial space to outcome distribution. *International Journal of Computers for Mathematical Learning,* **11**(1), 137–146.

Abrahamson, D. (2009) A student's synthesis of tacit and mathematical knowledge as a researcher's lens on bridging learning theory. *International Electronic Journal of Mathematics Education,* **4**(3), 195–226.

Adi, H., Karplus, R., Lawson, A. and Pulos, S. (1978) Intellectual development beyond elementary school UI: Correlational reasoning. *School Science and Mathematics*, 675–683.

Alloy, L.B. and Tabachnik, N. (1984) Assessment of co-variation by humans and animals: The joint influence of prior expectations and current situational information. *Psychological Review*, **91**, 112–149.

Atance, C.M. and Meltzoff, A.N. (2005) My future self: Young children's ability to anticipate and explain future states. *Cognitive Development*, **20**, 341–361.

Atance, C.M. and Meltzoff, A.N. (2006) Preschoolers current desires warp their choices for the future*. Psychological Science***, 17**, 583–587.

Barratt, B.B. (1975) Training and transfer in combinatorial problem solving: the development of formal reasoning during early adolescence. *Developmental Psychology,* **11**(6), 700–704.

Batanero, C. and Serrano, L. (1999) The meaning of randomness for secondary school students. *Journal for Research in Mathematics Education,* **30**, 558–567.

Batanero, C., Estepa, A., Godino, J.D. and Green, D.R. (1996) Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education,* **27**, 151–169.

Brown, M. (1981) Number operations. In Hart, K. (ed.), *Children's understanding of mathematics: 11–16* (pp. 23–47). Windsor: NFER–Nelson.

Bryant, P., Morgado, L. and Nunes, T. (1992) *Children's understanding of multiplication.* Paper presented at the Psychology of Mathematics Education, Tokyo, Japan.

Bullock, M. and Gelman, R. (1979) Preschool children's assumptions about cause and effect. *Child Development,* **50**, 89–96.

Carvalho, L.M.T.L. d. (2008) *O papel dos artefatos na construção de significados matemáticos por estudantes do ensino fundamental (The role artifacts play when elementary school students construct mathematical meanings).* Universidade Federal do Ceará, Fortaleza, Ceará, Brazil.

Castro, C.S., (1998) Teaching probability for conceptual change. *Educational Studies in Mathematics,* **35***,* 233–254.
Chandler, M. and Lalonde, C.E. (1994) Surprising, magical and miraculous turns of events: Children's reactions to violations of their early theories of mind and matter. *British Journal of Developmental Psychology,* **12**, 83–95.
Chapman, L.J. (1967) Illusory correlation in observational report. *Journal of Verbal Learning and Verbal Behavior,* **6**, 151–155.
Chapman, L.J. and Chapman, J.P. (1967) Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology,* **72**, 193–204.
Chapman, L.J. and Chapman, J.P. (1975) The basis of illusory correlation. *Journal of Abnormal Psychology,* **84**, 574–575.
Cheng, P.W. and Holyoak, K.J. (1985) Pragmatic reasoning schemas. *Cognitive Psychology,* **17***,* 391–416.
Cheng, P.W. Holyoak, K.L., Nisbett, R. and Oliver, L. (1986) Pragmatic versus syntactic approaches to training deductive reasoning. *Cognitive Psychology,* **18***,* 293–328.
Chernoff, E. (2009) Sample space partitions: An investigative lens. *Journal of Mathematical Behavior,* **28**, 19–29.
Chiesi, F. and Primi, C. (2009) Recency effects in primary-age children and college students. *International Electronic Journal of Mathematics Education,* **4**(3), 259–274.
Das Gupta, P. and Bryant, P.E. (1989) Young children's causal inferences. *Child Development,* **60**, 1138–1146.
Dawes, R.M. (2001) Probabilistic thinking. In Smelser, N.J. and Baltes, P.B. (eds.), *International Encyclopedia of the Social and Behavioral Sciences* (pp. 12082–12089). Amsterdam: Elsevier.
Denison, S., and Xu, F. (2009). Twelve to 14-month-old infants can predict single-event probability with large set sizes. *Developmental Science*, 1–6.
Desforges, A. and Desforges, G. (1980). Number-based strategies of sharing in young children. *Educational Studies,* **6**, 97–109.
Desli, D. (1994) *Proportional reasoning: the concept of half in part–part and part–whole situations*. University of London: Unpublished MSc thesis, Department of Child Development and Primary Education.
English, L. (1991) Young children's combinatoric strategies. *Educational Studies in Mathematics,* **22**, 451–474.
Evans, J.S.B.T. (1989) *Bias in human reasoning: Causes and consequences*. Hillsdale, NJ: Erlbaum.
Falk, R. (1991) Randomness – an ill-defined but much needed concept. *Journal of Behavioral Decision Making,* **4***,* 215–218*.*
Falk, R. and Konold, C. (1997) Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review,* **104**(2).
Falk, R. and Wilkening, F. (1998) Children's construction of fair chances: Adjusting probabilities. *Developmental Psychology,* **34**(6), 1340–357.
Falk, R., Falk, R. and Levin, I. (1980) A potential for learning probability in young children. *Educational Studies in Mathematics,* **11**, 181–204.
Fischbein, E. (1987) *Intuition in Science and Mathematics*. Dordrecht: Reidel.
Fischbein, E. and Gazit, A. (1984) Does the teaching of probability improve probabilistic intuitions? *Educational Studies in Mathematics,* **15**, 1–24.
Fischbein, E., Pampu, I. and Minzat, I. (1970) The effects of age and instruction on combinatory ability in children. *British Journal of Educational Psychology,* **40**, 261–270.

Frydman, O. and Bryant, P.E. (1988) Sharing and the understanding of number equivalence by young children. *Cognitive Development,* **3**, 323–339.

Gigerenzer, G. (2002) *Reckoning with Risk*. London: Penguin Books.

Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J. and Krüger, L. (1989) *The empire of chance: How probability changed science and everyday life*. Cambridge: Cambridge University Press.

Gilovich, T, Vallone, R. and Tversky, A. (1985) The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology,* **17***,* 295–314.

Girotto, V., Light, P. and Colbourn, C. (1988) Pragmatic schemas and conditional reasoning in children. *The Quarterly Journal of Experimental Psychology, Section A,* **40**, 469–482.

Gopnik, A., Glymour, C., Sobel, D.M., Schulz, L.E., Kushnir, T. and Danks, D. (2004) A theory of causal reasoning in children: Causal maps and Bayes nets. *Psychological Review,* **111**(1), 3–32.

Green, D.R. (1979) The chance and probability project. *Teaching Statistics,* **1**(3), 66–71.

Harris, P.L. and Kavanagh, R.D. (1993) Young children's understanding of pretense. *Monographs of the Society for Research in Child Development,* **58***(1-Serial no, 231)*, 1–92.

Hay, D.E., Caplan, M., Castle, J. and Stimson, C. (1991) Does sharing become increasingly 'rational' in the second year of life? *Developmental Psychology,* **27**, 987–993

Hoffrage, U. and Gigerenzer, G. (1998) Using natural frequencies to improve diagnostic inferences. *Academic Medicine*, **73**, 538–540.

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

Hoffrage, U., Lindsay, S., Hertwig, R. and Gigerenzer, G. (2000) Communicating statistical information. *Science,* **290**, 2261–2262.

Howe, C., Nunes, T. and Bryant, P. (2010) Rational number and proportional reasoning: Using intensive quantities to promote achievement in mathematics and science. *International Journal of Science and Mathematics Education,* published online 5 October 2010.

Inhelder, B. and Piaget, J. (1958) *The Growth of Logical Thinking From Childhood to Adolescence*. New York: Basic Books.

Jennings, D.L., Amabile, M.T. and Ross, L. (1982) Informal covariation assessment: Data-based versus theory-based judgments. In Kahneman, D., Slovic, P. and Tversky, A. (eds.), *Judgment Under Uncertainty: Heuristics and Biases* (pp. 211–230). New York: Cambridge University Press.

Johnson, C.N. and Harris, P.L. (1994) Magic: Special but not excluded. *British Journal of Developmental Psychology,* **12**, 35–51.

Jones, G.A., Langrall, C.W., Thornton, C.A. and Mogill, A.T. (1997) A framework for nurturing young children's thinking in probability. *Educational Studies in Mathematics,* **32**, 101–125.

Jones, G.A., Langrall, C.W., Thornton, C.A. and Mogill, A.T. (1999) Students' probabilistic thinking in instruction. *Educational Studies in Mathematics,* **30**(5), 487–519.

Kahneman, D. and Tversky, A. (1972) Subjective probability: A judgment of representativeness. *Cognitive Psychology,* **5**, 207–232.

Karplus, R., Adi, H. and Lawson, A.E. (1980) Intellectual development beyond elementary school VIII: Proportional, probabilistic, and correlational reasoning. *School Science and Mathematics,* **80**, 673–683.

Keren, G. (1984) On the importance of identifying the 'correct' problem space. *Cognition,* **16**, 121–128.

Konold, C., Harradine, A. and Kazak, S. (2007) Understanding distributions by modeling them. *International Journal of Computing and Mathematics Education,* **12**, 217–230.

Kuzmak, S. D. and Gelman, R. (1986) Young children's understanding of random phenomena. *Child Development,* **57**(3), 559–566.
Lawson, A. (1982) The relative responsiveness of concrete operational seventh grade and college students to science instruction, *Journal of Research in Science Teaching,* **19**(1), 63–77.
Lawson, A.E. and Snitgen, D.A. (1982) Teaching formal reasoning in a college biology course for preservice teachers. *Journal of Research in Science Teaching,* **19**, 233–248.
Lawson, A.E., Adi, H. and Karplus, R. (1979) Development of correlational reasoning in secondary schools: Do biology courses make a difference? *The American Biology Teacher,* **41**, 420–425.
Lax, E. (2004) *The mould in Dr. Florey's coat*. London: Little Brown.
Lecoutre, M. -P. (1992) Cognitive models and problem spaces in 'purely random' situations. *Educational Studies in Mathematics,* **23**(6), 557–568.
Lecoutre, M. -P. and Durand, R. -L. (1988) Jugements probabilistes et modèles cognitifs; etudes d'une situation aléatoire. *Educational Studies in Mathematics,* **19**, 357–368.
Martignon, L. and Krauss, S. (2009) Hands-on activity for 4th-graders: A toolbox for decision making and reckoning with risk. *International Electronic Journal of Mathematics Education,* **4**(3), 227–256.
Miller, K. (1984) The child as the measurer of all things: Measurement procedures and the development of quantitative concepts. In Sophian, C. (ed.), *Origins of Cognitive Skills* (pp.193–228). Hillsdale, NJ: Erlbaum.
Mlodinow, L. (2009) *The Drunkard's Walk*. London: Penguin.
Nesher, P. (1988) Multiplicative school word problems: Theoretical approaches and empirical findings. In Hiebert, J. and Behr, M. (eds.), *Number Concepts and Operations in the Middle Grades* (pp. 19–40). Hillsdale,NJ: Erlbaum.
Nickerson, R.S. (1998) Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology,* **2**, 175–220.
Nous, A. and Raven, R. (1973) The effects of a structured learning sequence on children's correlative thinking about biological phenomena. *Journal of Research in Science Teaching,* **10**(3), 251–255.
Nunes, T. and Bryant, P. (1996) *Children Doing Mathematics*. Oxford: Blackwell.
Nunes, T. and Bryant, P. (2009) *Children's Reading and Spelling: Beyond the First Steps*. Chichester, UK: Wiley-Blackwell.
Nunes, T., Bryant, P., and Hurry, J. (2004) *The role of awareness in teaching and learning literacy and numeracy in Key Stage 2*: Report presented to the ESRC-TLRP, Award #L139251015
Paparistodemou, E., Noss, R. and Pratt, D. (2008) The interplay between fairness and randomness in a spatial computer game. *International Journal for Computer Mathematics Learning,* **13**, 89–110.
Piaget, J. and Inhelder, B. (1975) *The Origin of the Idea of Chance in Children*. London: Routledge and Kegan Paul.
Pisa Consortium Deutschland (2004) *PISA 2003:Der Bildungsstand der Jugenlichen in Deutschland-Ergebnisse des zweiten internationalen Vergleichs.* Munster: Waxmann**.**
Polaki, M.V. (2002) Using instruction to identify key features of Basotho Elementary students' growth in probabilistic thinking. *Mathematical Thinking and Learning,* **4**(4), 285–313.
Pratt, D. (2000) Making sense of the totals of two dice. *Journal for Research in Mathematics Education,* **31**(5), 602–625.
Pratt, D. and Noss, R. (2002) The microevolution of mathematical knowledge: The case of randomness. *Journal of the Learning Sciences,* **11**(4), 453–488.

Robinson, J. (1968) *The Nature of Science Teaching*. Belmont: Wadsworth.
Rosengren, K.S. and Hickling, A.K. (1994) Seeing is believing: Children's explanations of commonplace imaginary and extraordinary transformations. *Child Development,* **65**, 1605–1626.
Ross, J.A. and Cousins, J.B. (1993) Enhancing secondary school students' acquisition of correlational reasoning skills, *Research in Science and Technological Education*, **11**(2), 191–205.
Russell, J., Alexis, D. and Clayton, N. (2010) Episodic future thinking in 3- to 5-year-old children: The ability to think of what will be needed from a different point of view. *Cognition*, **114**, 56–71.
Scholz, R.W. (1991) Psychological research in probabilistic understanding. In Kapadia, R. and Borovnick, M. (eds.), *Chance Encounters: Probability in Education* (pp. 213–249). Dordrecht (The Netherlands): Kluwer.
Schultz, T. (1982) Rules of causal attribution. *Monographs of the Society for Research in Child Development,* **47**.
Shtulman, A. (2009) The development of possibility judgments within and across domains. *Cognitive Development,* **24**, 293–309.
Shtulman, A., and Carey, S. (2007) Impossible or improbable? How children reason about the possibility of extraordinary claims. *Child Development,* **78**, 1015–1032.
Smedslund, J. (1963) The concept of correlation in adults. *Scandinavian Journal of Psychology,* **4**, 165–173.
Sobel, D.M. and Kirkham, N.Z. (2007) Bayes nets and babies: infants' developing knowledge of statistical reasoning and their representation of causal knowledge. *Developmental Science,* **10**, 298–306.
Sobel, D.M., Tenenbaum, J.B. and Gopnik, A. (2004) Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science,* **28***,* 303–333.
Spinillo, A. and Bryant, P. (1991) Children's proportional judgements: The importance of 'half'. *Child Development,* **5**, 427–440.
Spinillo, A. and Bryant, P. (1999) Proportional reasoning in young children: part–part comparisons about continuous and discontinuous quantity. *Mathematical Cognition,* **5**, 181–197.
Squire, S. and Bryant, P. (2002) From sharing to dividing: Young children's understanding of division. *Developmental Science,* **5**, 452–466.
Subbotsky, E.V. (2004) Magical thinking in judgements about causation: Can anomalous phenomena affect ontological beliefs in children and adults? *British Journal of Develomental Psychology,* **22**, 123–152.
Suddendorf, T. and Busby, J. (2005) Making decisions with the future in mind: Developmental and comparative identification of mental time travel. *Learning and Motivation*, **36**, 110–125.
Suddendorf, T. and Corballis, M.C. (1997) Mental time travel and the evolution of the human mind. *Genetic, Social, and General Psychology Monographs,* **123**, 133–167.
Suddendorf, T. and Corballis, M.C. (2007) The evolution of foresight: What is mental time travel, and is unique to humans? *Behavioural and Brain Sciences,* **30**, 299–351.
Teglas, E., Girotto, V., Gonzales, M. and Bonatti, L. (2007) Intuitions of probabilities shape expectations about the future at 12 months and beyond. *Procedings of the National Academy of Sciences***, 104**(48), 19156–19159.
Tobin, K.G. and Capie, W. (1981) The development and validation of a group test of logical thinking. *Educational and Psychological Measurement,* **41**, 413–423.

Tversky, A. and Kahneman, D. (1971) Belief in the law of small numbers. *Psychological Bulletin,* **76,** 105–110.

Tversky, A. and Kahneman, D. (1974) Judgment under uncertainty: Heuristics and biases. *Science,* **185**, 1124–31.

Tversky, A. and Kahneman, D. (1982) Evidential impact of base rates. In Kahneman, D., Slovic, P. and Tversky, A. (eds.), *Judgement Under Uncertainty: Heuristics and Biases* (pp. 153–160). Cambridge: Cambridge Univerity Press.

Van Dooren, W., Bock, D., Depaepe, F., Janssens, D. and Verschaffel, L. (2003) The illusion of linearity: The evidence towards probabilistic reasoning. *Educational Studies in Mathematics,* **53**, 113–138.

Vass, E. Schiller, D. and Nappi, A.J. (2000) The effects of instructional intervention on improving proportional, probabilistic, and correlational reasoning skills among undergraduate education majors. *Journal of Research in Science Teaching,* **37**, 981–995.

Wason, P.C. (1968) Reasoning about a rule. *Quarterly Journal* of *Experimental Psychology***, 20**, 273–281.

Watson, A., Jones, K. and Pratt, D. (in press)*. Key Ideas in Teaching Mathematics: Research-based guidance for ages 9–19*. Oxford: Oxford University Press. Scheduled for publication in 2013.

Watson, J.M. and Moritz, J.B. (2003) Fairness of dice: A longitudinal study of students' beliefs and strategies for making judgments. *Journal for Research in Mathematics Education,* **34(**4), 270–304.

Woolley, J. and Cox, V. (2007) Development of beliefs about storybook reality. *Developmental Science,* **10**, 681–693.

Xu, F. and Tenenbaum. J.B. (2007) Word learning as Bayesian inference. *Psychological Review,* **114***,* 245–272.

Xu, F. and Denison, S. (2009) Statistical inference and sensitivity to sampling in 11-month-old infants. *Cognition,* **112**, 97–104.

Xu, F. and Garcia, V. (2008) Intuitive statistics by 8-month-olds. *Proceedings of the National Academy of Sciences,* **105**, 5012–5015.

Zhu, L. and Gigerenzer, G. (2006) Children can solve Bayesian problems: The role of representation in mental computation. *Cognition,* **98**, 287–306.