In F.S. Gordon & S.P. Gordon (Eds.), Statistics for the twenty-first century (pp. 151 - 164). Mathematical Association of America. (1992)

SRRI #242

The Psychology of Learning Probability[1]

Ruma Falk[2] and Clifford Konold

Department of Psychology and School of Education, Hebrew University of Jerusalem

Scientific Reasoning Research Institute, University of Massachusetts, Amherst

Today our vision of the world is permeated by probability, while in 1800 it was not. Probability is the great philosophical success story of the period.

Hacking, in [8]

1. Introduction

Probability is too frequently regarded as a subsidiary topic of statistics. Consistent with this view, statistics educators often teach the minimum amount of probability they regard as sufficient for learning statistics. We regard this as a grave mistake. *Probability is a way of thinking*. It should be learned for its own sake. In this century probability has become an integral component of virtually every area of thought. We expect that understanding probability will be as important in the 21st century as mastering elementary arithmetic is in the present century.

The term "probabilistic revolution" [8 and 9] broadly suggests a shift in world-view from a deterministic description of reality, phrased in terms of universal laws of stern necessity, to one in which probabilistic ideas have become central and indispensable. Concepts of uncertainty are introduced into science partly because *we are ignorant* of the multiplicity of variables affecting our data, and because there is error in our measurements. Hence, even those who believe in the ultimate determinism of nature, nevertheless use probabilistic language to account for human shortcomings. The revolution goes further, however, viewing chance as an irreducible part of natural phenomena. That conception is typically represented by the *inherently indeterminate* view of quantum physics. An equally dramatic example of a probabilistic revolution that has changed our thinking about our own existence is found in evolutionary biology.

In learning probability, we believe the student must undergo a similar revolution in his or her own thinking. There are good reasons to expect that this breakthrough cannot be easy. There might have been a firm psychological basis for the historical persistence of the deter-

[1]Received: November 1990; Revised: August 1991

[2]This article was written while Ruma Falk was on sabbatical leave at the Department of Psychology, University of Massachusetts, Amherst. Preparation of the chapter was supported in part by the Sturman Center for Human Development, the Hebrew University, Jerusalem, and by the National Science Foundation, Grant MDR-8954626. We wish to thank Raphael Falk for his helpful comments.

ministic outlook in science. As we shall show, students (and some teachers) are apriori widely inclined to think in causal, deterministic terms and to overlook the effects of chance.

In our statistical age, *public* discourse is cast in the language of statistics and probability. However, our *private* musings and observations may still be subject to the old deterministic habits of thought. Whatever our psychological leanings and untutored intuitions, "for certain public purposes we shift our gaze from individuals to averages, from deep-felt certainties to probabilities, from impressions to numbers" [4, p. 291]. The teaching of probability should strive not only to provide students with the framework necessary for carrying on such a public discourse, but also to transform their private "musings."

We first focus on some major psychological obstacles inherent in making the switch to probabilistic reasoning. Next, we illustrate a few prevalent misconceptions of chance, adding a caveat for the teacher not to overdo confronting students with their own erroneous intuitions. Finally, we advocate starting the process of probabilistic education by building on the firm basis of students' sound intuitions, endeavoring to strengthen students' conceptual grip of probability theory without getting too technical. We address issues that should be relevant to a wide range of levels of introductory courses.

2. The Search for Certainty

All that happens around children is interwoven with probabilistic elements, and all they hear and say, with logical ones.

T. Varga

2.1 Decisions versus outcomes

Some years ago the first author devised a probability game in which young children's ability to compare probabilities [3] could be practiced: each of two players is required to choose which of two discs to spin before moving on a game board. Each of the discs is divided into sectors of two colors, one favorable and one unfavorable. The discs are divided in different proportions, and it is to a player's advantage to select and spin the disc with the greater probability of a favorable outcome. The game was sharply criticized by parents and educators as being 'uneducational.' They objected to the notion of a game in which one might make a correct choice (of the disc with a higher probability of success) and yet obtain an unfavorable outcome, while on the other hand, an incorrect decision may be rewarded. Obviously, they wished for a consistent, 'just' system. Implied in their criticism was the expectation that good decisions would always be reinforced, while bad ones would never be. Apparently, the two concepts of a *correct choice* and a *favorable outcome* were not clearly differentiated in these critics' minds.

Learning and acquiring knowledge are typically associated with increasing certainty and dispelling doubt. Hence, the primary objective in teaching probability is to reconcile students' habitual search for (certain) knowledge with the understanding that learning probability will not enable us to predict the immediate outcome of a random process. Even the most probable event, which is expected to occur most frequently in the long run, is not guaranteed to occur on the next opportunity.

2.2 Underrating the role of chance

The tendency of students of all ages either to reject uncertainty altogether or to underestimate its impact has been documented in a variety of educational and psychological studies [e.g., 3]. Young children of about age 6 said of a spinning top with a 2/3 probability of success "Now I am sure to win." Probabilities greater than 1/2 were all "sure to win," while those below 1/2 were "sure to lose." Children playing a game of chance in which they repeatedly observe the different possible outcomes, often change their minds about their correct choices when they are not promptly rewarded. They also stick to incorrect choices which happen to result in a favorable outcome. Similarly, some adults [7] consider a weather forecast stating that the probability of rain on a particular day is 70% to be wrong if it does not rain that day.

Many of the biases and fallacies found in the research on judgment under uncertainty [6] result from employing various heuristics which may roughly be characterized as minimizing the role of chance. There is widespread confusion between the terms 'representative sample' and 'random sample.' People tend to identify a random sample with a *true cross section* of the population, neglecting the ever-present variation due to chance sampling. A variety of studies conducted in the last decades offer a fairly coherent picture of people's beliefs about randomness: disregarding chance fluctuations, people expect population proportions to be represented not only globally in the entire random sequence, but also locally in each of its parts. They regard chance as a self-correcting mechanism in which a deviation in one direction is *promptly* balanced by a deviation in the opposite direction. That view clearly ignores the most significant characteristic of chance, namely, its *blindness*. Unlike the human observer, a chance process does not review its productions and feels no obligation to restore equilibrium in a sample of any size.

Nonregressive predictions, typically obtained in judgment research [6], can also be construed as a form of denial of chance's contribution. To expect sons' heights to be as extreme as those of their fathers is to ignore the imperfect validity of father's height as a predictor. The overlooked variables, such as mother's genes and pre- and post-natal environmental factors, are either weakly correlated with, or completely independent of, father's height. It is because these other multiple factors are nonsystematic, or *random*, that we get regression toward the population mean. An extreme reliance on the input as predictor is, in fact, a typical manifestation of shunning chance.

One of the first goals in teaching probability should be to help the student recognize the fact that chance cannot be 'driven out of the system.' No matter what progress we make in our study of probability or of the phenomena of interest, uncertainty cannot be eliminated. The best we can hope for is to sharpen our tools for accurately quantifying our long-run expectations (or our degrees of uncertainty, depending on the interpretation of probability that we endorse).

2.3 Short versus long run

> *When randomness is present, probability answers the question, "How often in the long run?" and expected value answers the question, "How much in the long run?"*
>
> *D. S. Moore*

Probability is one of many mathematical and physical concepts that have been studied by cognitive-developmental psychologists using the binary-choice paradigm. In these experiments, two stimuli are presented in each trial, and the subject is required to judge in which of them the studied variable assumes a greater value. Concepts studied using this paradigm include number of objects, liquid quantity, equilibrium in a balance-scale apparatus, proportionality of orange-juice concentration, and sweetness of sugar solutions. Piaget and Inhelder [11] were first to present binary comparisons of probabilities to children, and many studies have since introduced modifications and improvements to this paradigm. One significant feature, however (the educational implications of which are far-reaching), distinguishes probability from the group of studies of other mathematical-physical concepts: in the latter, a correct judgment is invariably accompanied by positive feedback. The correctness of choice between the two sides of the balance scale, or between two cups of sugar solution, can be promptly verified by a single experimental trial. In the probabilistic situation, however, the role of immediate experimental feedback is equivocal.

If success is not guaranteed even though one chooses the alternative with the highest probability of success and, moreover, choosing the option with a lower success-probability may nevertheless prove successful, in what sense is the former choice correct and the latter incorrect? The answer is to be found in the *long run*. According to the frequency interpretation of probability, the probability of success is the limiting proportion of successful trials when the number of repeated trials grows indefinitely. A test of the correctness-of-choice between two urns with different proportions of beads of the winning color requires multiple draws. Furthermore, regardless of the sample size, the conclusion is still an inference with some, though small, degree of uncertainty.[3]

Suppose that, in accord with the short-run expectations of the critics of the game described above, we arrange that a choice with 75% probability of success is 100% rewarded. In that case we would be teaching children something incorrect! The real (uncertain) world is so structured that even wise decisions are only probabilistically reinforced. It is only in the long run that we should expect good decisions to pay off. Thus, choice of the disc constructed of the highest proportion of the winning color is justified, not because it ensures an immediate reward, but because it will turn up most favorably over *many repetitions*. Simple as that statement must seem, it is not easily internalized and it defies many students' intuition.

The concept of probability can be decomposed into two subconcepts: proportion and chance. It is because of chance that a correct decision proves true only in the long run. The distinction between a correct probabilistic choice and a favorable outcome is apparently quite elusive since, in most of the studies of choice between probabilities, investigators have overlooked the distinction, analyzing the task as if it involved only comparison of proportions.

The complicated status of feedback as a vehicle of instruction poses a serious challenge to the teacher of probability. In discouraging students from regarding *one* experimental result (despite its evidential value) as an ultimate criterion of correctness, we do not want to diminish

---

[3]It should be noted that on a deeper level of analysis (in accord with the probabilistic world-view) the difference in the role of feedback between 'deterministic' experiments and a random experiment can be construed as just a difference in degree. It is not absolutely certain that a sip from the cup with the stronger sugar solution will indeed taste sweeter. Even when the solutions have been properly stirred, there is always a non-zero (though practically negligible) probability that a sample of that cup may taste less sweet than a sample of the other cup. It is because of the gigantic number of particles in any small sample of the sugar solution that success in a single tasting from the correct cup is virtually guaranteed. This is simply a manifestation of the "law of large numbers." Were we to devise a mechanism of drawing single molecules from the cup, the situation would reduce to that of drawing beads from an urn.

in their minds the importance of observing relevant data; we simply want to shift their perspective from the very short to the long run.

2.4 Prediction versus quantification

> *I don't believe in probability, because even if there is a 20% chance, it could happen. Even 1%, it could happen. I don't believe in probability.*
>
> *Student Quotation, July, 1987*

While most students entering their first course in probability do not come with an alternative discipline of thought, they do speak a language different from that of the teacher (or at least that part of the teacher who stands ready with an urn and a pocket full of formulae). This is because students, as inhabitants of an uncertain world, have learned to think and converse in everyday language about 'chance,' 'probability,' 'luck,' 'randomness,' and have developed a rich vocabulary with which to communicate degrees of belief from 'I'm certain' to 'it's impossible.' Underlying these terms exists a perspective on uncertainty that is, at points, at odds with formal theory.

A very basic difference between formal and informal views of probability concerns the perceived objective in reasoning about uncertainty. Formal probability is mostly concerned with deriving measures of uncertainty, or answering the question '*How often* will event A occur *in the long run?*' On the other hand, what most people want is to *predict* what will occur in *a single instance* — to answer the question '*Will* A occur or not?'

Konold [7] has referred to this latter perspective as the "outcome approach." People reasoning via the outcome approach tend to interpret a request for a probability of some event as a request to predict whether or not that event will occur on the *next* trial. For example, subjects were given an irregularly shaped bone that could rest (with unequal probabilities) on six different surfaces [7]. They were asked, "If you were to roll this, which side do you think would most likely land upright?" After answering this question, subjects rolled the bone. Several spontaneously judged their answer as being wrong (or right) after that single roll. And as mentioned previously, subjects evaluated the forecast "70% chance of rain tomorrow" as being incorrect based on the information that it did not rain. Furthermore, given records of what happened on 10 days for which a 70% forecast of rain had been made on each day, subjects judged the forecaster as performing suboptimally when, in fact, it rained on 70% of those days. According to these subjects, a perfect performance would have entailed rain on all 10 days.

Understanding that many students view the objective of probability as making predictions can help the teacher of probability make sense of what otherwise are incomprehensible statements. The statement at the beginning of this section was made by a student about half-way through a two-week workshop on probability. Our interpretation of this student's claim is that she is in the process of giving up her outcome orientation. She is questioning her ability to make accurate single-trial predictions. Note that if the phrase 'making single-trial predictions' is substituted for the word 'probability,' the claim is understandable and normative. What this student has not yet come to realize, however, is that the word 'probability' refers to something other than single-trial predictions.

Given the desire for predictions, outcome-oriented subjects translate probability values into yes/no decisions. A value of 50% is interpreted as total lack of knowledge about the outcome, leaving one without justification for making a prediction. For example, rather than interpreting the 50% probability of heads in flipping a fair coin as the expected percentage of heads in the long run, many subjects interpret it as "You can't tell...it could be anything." Similarly, a forecast of 50% chance of rain is taken as an admission of total ignorance: "If he said 50/50 chance, I'd kind of think that was strange ... that he didn't really know what he was talking about ... ." ([7] p. 68).

Probability values sufficiently above or below 50% are coded as "yes" or "no" predictions, respectively, as illustrated below. (*I* and *S* stand for Interviewer and Subject, respectively.)

> I: What does the number, in this case the 70%, tell you?
>
> S: Well, it tells me that it's over 50%, and so, that's the first thing I think of. And, well, I think of the half-way mark between 50% and say 100% to be like, well, 75%. And it's almost that, and I think that's a pretty good chance that there'll be rain. ([7] p. 66.)

Konold [7] also found a correlation between the tendency to interpret questions about probability as requests for single-trial predictions and a preference for analyzing probabilistic situations from a causal perspective. Indeed, the belief that one could successfully predict single instances would seem to require an understanding of the causal mechanisms involved. One component of this causal orientation is the preference for basing predictions on information that seems causally related to the outcome of interest rather than on frequency information. Subjects who tended to evaluate predictions as right or wrong after a single trial also tended to give little weight to the results of rolling the bone. For example, after rolling the bone 7 times and getting 3 Ds, 2 Cs and 2 Bs, the subject below judged C to be the most likely based on an analysis of the 'shape' of the bone. She was then asked:

> I: Would rolling that a lot more times help you, in any way, decide which side is most likely to come up?
>
> S: I don't think so.

She was then given a list showing the results of 1,000 rolls.

> I: Would you feel safe in concluding, looking at this table, that D is, in fact, more likely than B? [Frequencies were A-50, B-279, C-244, D-375, E-52, F-0]
>
> S: I don't know. Not really. I think it could just have something to do with your luck, or your chance. If I did the same number of rolls, I don't know if it would come out the same.
>
> I: What is your belief about which side is most likely to come up if you rolled it?
>
> S: I have to look at this [bone] again. It changed shape in the past few minutes. [Laughter] I still think C.

In addition to underweighting frequency data, a few subjects suggested that the number in the statement "70% chance of rain" was a measure of some factor causally related to rain -- "Maybe it means that there's 70% humidity in the air... ."

Most instruction in probability ignores problems associated with interpretation, so that disagreements are allowed to remain unexamined. These ultimately reveal themselves, however, in students' inability to transfer or to show other signs of understanding. Students who reason according to the outcome approach cannot be treated as if they are suffering from a simple misunderstanding. It does not do, for example, for the teacher to keep repeating, "What I mean by probability is... ." Each student will have to struggle with various contradictions that the outcome approach (and judgment heuristics) entails before they are able to adopt a more normative view.

The difficulties for students of probability are not over, however, once they have accepted uncertainty and adopted the convention of representing this uncertainty quantitatively. The problems we discuss below are characteristic of those that continue to plague students even after they have gotten past the more rudimentary aspects that we have discussed to this point.

3. Difficulties in Applying the Probabilistic Model

Probability is especially rich in counterintuitive examples, which often entail fallacies and paradoxical conclusions. Some of these examples played an important role in the development of probability theory. Students may likewise benefit from comparing their intuitions concerning puzzles and paradoxes with normative solutions. This activity requires increased awareness of one's own thought processes. Knowledge of one's own thinking (metacognition) is no less important than learning the right solution, and reflective thinking is a vital step toward achieving abstract mathematical ability. A paradox usually triggers a conflict. Such conflicts may encourage students to critically examine their intuitive theories. In parallel to the historical development of the theory, this examination might reveal to students deficiencies in their understanding and so promote the development of normative concepts. In addition, many difficulties in probability arise from ambiguities and complexities involved in translating real-world situations into formal ones. Solving an intriguing puzzle requires an unequivocal definition of the problem at hand. The problem is often solved once the assumptions involved in modeling are made explicit. These possibilities will be demonstrated via a few examples.

3.1 The gambler's fallacy

> *Many people seem to believe that there is something akin to the law of atonement in probability. If a coin thought to have no built-in bias produces nine heads in a row, they reason, one would be wise to bet on tails the next flip. After all, if the ratio of tails to total flips is ½ in the long run, then, in the aforementioned long run, there should be as many tails as heads and that means tails have some catching up to do. The fallacy in this line of reasoning lies in the fact that the coin has no memory nor conscience. The coin simply isn't aware that tails are lagging behind heads.*
>
> *S. K. Campbell*

Events in a random sequence are statistically independent of each other. The concept of *statistical independence*, though easy to formally define via simple equalities of probabilities, is not easily internalized. Observers find it difficult to avoid the perception of interdependence among events that are in fact unrelated. That fallacy is revealed not only in the gambling setting, but in diverse tasks, such as generation and judgment of randomness and assessment of likelihoods. The gambler's fallacy appears in varied disguises. People typically regard long runs that turn up by chance as non-random occurrences. Thus, when they observe sequences of hits and misses in a basketball game, they attribute the random runs, which *seem* too long for the production of chance, to a mysterious, non-random agent [5]. This agent has been dubbed the 'hot hand' in the case of long runs of hits. In complete analogy, the attention of casino gamblers is drawn to random runs of wins or losses that are subjectively too long. They similarly invent a non-chance agent, 'luck,' to account for these apparent deviations from randomness [12].

Students should be encouraged to view statistical independence from different perspectives in the hope that they will assimilate its meaning to the point of eliminating these notorious biases. But overcoming these biases will not be easy; they are quite compelling. Nor are they limited to naive subjects; even experts can fall prey to the fallacies. Consider the following problem,[4] which is addressed to a large class of students.

> Suppose each of you writes down on a note how many brothers and how many sisters you have. We collect the notes and separate them into those written by women and those written by men. Should we expect men to have more sisters than women have? and would men have more sisters or more brothers? or are no differences expected?

Despite agreeing to assume equal probabilities for male and female births and independence among births, most people expect men to have both more sisters than women have and more sisters than brothers. They reason that because, 'on the average,' families have an equal number of sons and daughters, the set of siblings of men, who are themselves not counted, should comprise an excess of women (sisters). This compelling reasoning is nevertheless wrong! Men are expected to have the same number of sisters as women have, and the same number of sisters as brothers.

Overcoming the gambler's fallacy in this case requires a full realization of the significance of statistical independence. Asking a random child from a family of n children how many brothers and how many sisters he or she has, is equivalent (mathematically) to picking a random family with n-1 children and asking how many sons and how many daughters it has. In the latter case we expect, however, the same number of sons and daughters. Hence, irrespective of whether the questioned persons are males or females, they would have, on the average, the same numbers of brothers and sisters.

Even though the above reasoning seems right, some doubt may linger in students' minds. The feeling may persist that men must have more sisters since the men are not counted in the statistics of their families. That doubt may be enhanced by the apparent similarity to sampling *without* replacement from an urn with half Ms and half Fs. It is important, therefore, to dwell further on this problem until one clearly sees that asking a man about the sexes of his siblings is -- statistically speaking -- identical to asking a woman the same question. This is the essence of the concept of statistical independence: the probabilities of male or female births in the respondent's family are *not affected* by knowledge of the respondent's sex.

Some students may nevertheless experience some conflict between the long-run expectation of equal proportions of males and females and the exclusion of the respondents from consideration in a finite population. The following is a representative argument that was given

[4]We advise the readers to try to solve this, as well as the other problems, before reading our exposition of the solution.

by a male student in one of our classes: "Suppose there were 1,000 men in our class. Consider all 1,000 of us together with our brothers and sisters. That large group should comprise approximately equal proportions of males and females. Now, exclude us, 1,000 men, and there would remain more sisters than brothers." This reasoning fails to take into account the fact that all-daughter families are not represented among those 1,000 families, and, moreover, that families with many sons are overrepresented in that large group. The group will therefore contain an excess of males.[5] Only after removing the original 1,000 men should the expected number of males (brothers) equal that of females (sisters).

3.2 Overlooking the chance mechanism

Problems about sex distributions of children in families provide a useful context for applying basic probability theorems. The example below involves only two children, yet it is challenging even to those who accept the conventional assumptions concerning sex determination.

> Mrs. F. is known to be the mother of two children. You meet her in town with a boy whom she introduces as her son. What is the probability that Mrs. F. has two sons? One possible answer is that the probability is 1/2: you now have seen one boy, and hence the question is whether the other child is a male. Another possibility is that the probability is 1/3: you learned that Mrs. F. has 'at least one boy,' and hence three equiprobable family structures are possible (BB, BG, GB), of which the target event (BB) is but one.

Both arguments make sense. No further discussion can decide between the two, because each may be true depending on the assumptions one makes. The explication of the assumptions, however, is missing in the problem's statement. The solution should depend on the exact method by which the observation has been obtained, namely, on the chance mechanism that has generated the datum. If the woman in question typically chooses at random one of her two children to accompany her on her outings, then your observation is 'a randomly selected child of the two-children-family was found to be a boy.' In such a case, a family with two sons is twice as likely to yield that observation than a mixed family. A simple Bayesian calculation shows that the posterior probability that the woman has two sons is 1/2 (see [1]). If, however, being a part of a male-chauvinistic society, she always prefers taking a son along with her, then your observation becomes 'the family has at least one son' and the three families of that kind are equally likely to bring about the above meeting. Consequently, the probability of two sons should be one third.

There is obviously no point in arguing about the correct answer to a problem that is under-defined. We could only *favor* one or the other of the assumptions as more reasonable. Lacking any other information, the most sensible assumption would be that chance determines the child to accompany the mother on a given trip. Indeed, most real-life situations, from which you learn that a given woman has a son, are similarly structured: you overhear the woman mentioning a son, or you pay a visit and catch a glimpse of a boy. In each of these cases, the greater the proportion of sons among the children in the family, the more likely is a chance meeting with one of its sons. Inventing a scenario in which such an encounter is equally likely

[5] The example 'inadvertently' provides an interesting opportunity to point out the sampling bias introduced by sampling families via their sons.

under all three family compositions -- BB, BG, and GB -- is considerably more difficult. The answer 1/2 is thus favored over 1/3 only in the sense of regarding the assumptions it is based on to be more realistic.

Textbooks of probability theory often begin with the concept of a *statistical experiment*, namely, an idealized, well-defined chance process whose elementary outcomes comprise the sample space. Indeed, probability theory is concerned with the outcomes of statistical experiments. As demonstrated in the analysis of the above family problem, the probabilistic conclusions depend on the exact definition of that random process. Describing the habits of the mother in choosing the child to accompany her to town amounts to modeling her behavior as formally equivalent to one urn model or another. The statistical experiment should clarify whether one child is randomly sampled from a (random) two-children family, or whether one family is randomly sampled from two-children families having at least one son.

Assumptions play a central role in reasoning and problem solving in all areas. Challenging one's assumptions is especially important when solving probability problems because of the subtle intricacies of modeling real-life situations. The psychologists Nisbett and Ross advise teachers to offer useful slogans for didactic purposes. One of their slogans, *Which hat did you draw that sample out of?* may be instrumental in alerting the students to reflect on the chance mechanism underlying the problem. A more general reminder would be, *Make implicit assumptions explicit*! Keeping that advice in mind, let us turn to another problem.

The "Flaws, Fallacies, and Flimflam" column of the *College Mathematics Journal* (January, 1990, p. 35) offered the following teaser from the folklore of probability (suggested by R. Guy):

> *Where the Grass is Greener.* There are two cards on the table. One of them has written on it a positive number; the other, half that number. One of the cards, selected by a coin flip, is revealed to you. You may get in dollars either the number on this card or the number on the other card. Which should you choose?
>
> Suppose that the number revealed to you is A. Then the other card has the number $2A$ or $0.5A$, each with equal probability. If you stick with the card shown, your expected winnings are A. If you switch, your expected winnings are $0.5(2A) + 0.5(0.5A) = 1.25A$. Thus, you should always select the card other than the one revealed to you.

Moreover, even without having seen the number on the first card you should switch, but then you should also switch back, and so on. One is caught in a never-ending pendulum swing. But is the argument indeed valid for *any positive A*? Let us examine what we know about the positive numbers written on these two cards. How did these numbers come about? We are uncertain about their values, but we ought to know the (chance) mechanism that has generated them. Nothing, however, is said about the selection process in the problem.

Suppose you were told that a person had drawn at random one of the integers 1, 2, 3, ..., 10; the outcome of that draw was written on one card, and twice that number on the other card. Now, considering that experimental procedure, if the number revealed to you were, say, 18, you would know that you should stick with it. If, however, you were shown a 7, you would switch. Different considerations could be applied, depending on the number revealed to you, and on the description of the number-generating process. Lacking any specification of the chance procedure in the problem statement, your considerations must be based on some *assumed* distribution of the pairs of numbers.

The crux of the puzzle was based on the assertion that given any $A > 0$, the probabilities that the other card bears either $0.5A$ or $2A$ are equal. The one assumption that is compatible with that requirement for *each* positive A, at least in the discrete case, is that the first number was randomly drawn from the set of all, say, integer powers of 2, and then it was doubled. Put differently, your expected winnings upon switching would be $1.25A$, *for every* A, only if you assume the first number was randomly drawn from an infinite uniform distribution. A uniform distribution, however, *cannot* be infinite. If a discrete random variable is uniformly distributed, it can assume only a finite number of values. On first reading, the reader is not aware of the fact that he or she is accepting an impossible statistical experiment. The paradoxical conclusion can, therefore, be traced to contradictory implicit assumptions.

Is there a chance set-up according to which it is always reasonable to switch? Suppose the game is conducted so that you are first shown some positive number W. Then, out of your sight, somebody flips a coin to decide whether the other card will bear the number $0.5W$ or $2W$. You are now to decide whether to accept W dollars as your winning or to switch. In this case, you would do better to switch. The lesson from this can be summed up by the admonition, *Inquire about the chance mechanism!*

4. Probabilistic Reasoning as an Extension of Commonsense Thinking

4.1 Grounding instruction on students' valid intuitions

Puzzles are often piquant. They may challenge students and capture their interest. It is tempting to bring some of the more devious problems to the classroom to demonstrate to students their erroneous tendencies and perhaps enlighten them. However, if a teacher persists in pointing out to students how prone they are to inferential errors, they may become so convinced of their incapacities that they despair of ever mastering more appropriate techniques. A balance needs to be struck between illustrating some misconceptions and biases and reassuring students of their existing capacity. Furthermore, it seems reasonable to begin instruction in probability by building on students' sound intuitions. Valid probabilistic intuitions are not hard to come by. Despite the abundance of studies describing people's inferential biases and shortcomings, many of the rules prescribed by probability theory are compatible with commonsense. Indeed, much of what is known as Bayesian inference is intuitively sensible. Everybody will agree, for example, that when a sick child develops a rash, the probability of measles should increase relative to that of flu. That is so because, no matter what the prior probabilities of the two diseases, a rash is more likely assuming measles than assuming flu.

In probability, as in physics, "not all preconceptions are misconceptions." That expression is borrowed from the title of a paper by Clement, Brown, and Zietsman [2]. They propose that in teaching physics it is desirable to ground new material in students' intuitions that are in agreement with accepted theory. Likewise, in teaching probability, one expedient strategy is to offer nontrivial probability problems for which students can guess whether the answer is greater or smaller than a given number. This would allow the student to experience the satisfaction of having the prediction borne out by the results of the probabilistic calculation. They may realize that commonsense can still be a good guide though it should be exercised with caution.

4.2 Bayesian inference and commonsense

Many people's sound intuitions in learning from experience and revising their beliefs are consistent with Bayesian analysis. Furthermore, in specific limiting cases the results of Bayesian computations coincide with those of deductive inference. Thus, Bayesian reasoning can be viewed as an extension of logical inference into the gray area of uncertainty. A clear and comprehensive exposition of Bayesian statistics can be found in [10].

We give one simple example that can be presented in introductory courses and may appeal as commonsensical even to children.

> A man was arrested as a suspect of murder. The investigating officer summed up his impressions of the suspect and all the relevant information at his disposal and arrived at the conclusion that the suspect's probability of guilt was .60.

a. As the investigation went on, it was found (beyond any doubt) that the murderer's blood-type was O. The relative frequency of blood-type O in that population is .33 (i.e., this is the probability that a 'random person' in that population has blood-type O). The suspect's blood was tested and found to be of type O. What is the posterior probability of the suspect's guilt (from the officer's point of view) considering all the data?

b. Suppose both the murderer's and the suspect's blood-types were found (with certainty) to be A. The relative frequency of blood-type A in the population is .42. How would the posterior probability in that case compare with the same probability in question (a)?

c. Suppose everything is as in question (a) above. The murderer's blood-type is O, but the suspect's blood-type is B. What is the suspect's posterior probability of guilt?

Let us denote the event "the suspect is guilty" by G. Guessing the direction of change in the probability of G in the three cases would pose no difficulty to most students. It would be gratifying, however, to see that formal computations yield the same conclusions.

a. Let "O" denote the evidence that the suspect's blood-type was found to be O. We know that $P(O|G) = 1$ and $P(O|\bar{G}) = .33$, and we are interested in the probability of guilt, given that evidence, namely, in $P(G|O)$. Now, employing Bayes' formula, one obtains:

$$P(G|O) = \frac{P(O|G)P(G)}{P(O|G)P(G) + P(O|\bar{G})P(\bar{G})}$$

$$= \frac{1 \times .60}{1 \times .60 + .33 \times .40} = .82$$

verifying the intuition that the result of the blood test should be incriminating for the suspect.

b. Commonsense clearly indicates that blood-type A would make the probability of the suspect being guilty smaller since there are more people with that blood type. When substituting A for O as the evidence in Bayes' formula, the only change would be to replace .33 in the denominator by .42, thus increasing the denominator and decreasing the resulting probability of guilt. Students may speculate about the limiting cases where either 100% of the population share the murderer's and suspect's blood-type, leaving the suspect's probability of guilt unchanged, or, there is known to be just one person in the population with the blood-type found to be the murderer's and the suspect's, causing our suspicion to jump to certainty. An important insight that students ought to distill from these various examples is that the lower the rate of the blood-type shared by the murderer and suspect, the more incriminating the evidence.

c. If the murderer's blood-type is O and that of the suspect is B, the latter is obviously exonerated. Formally, one has to replace $P(O|G) = 1$ in the formula by $P(B|G) = 0$, resulting in $P(G|B) = 0$. That logical result is obtained as an extreme case of the probabilistic model.

In comparing parts (a) and (c), we note that although $P(O|G) = 1$, observing O [in (a)] did not prove G; it only raised its probability (from .60 to .82). On the other hand, because $P(B|G) = 0$, observing B [in (c)] disproved G, thus proving the suspect's innocence. This is a demonstration of the conclusiveness of negative evidence compared with the relative corroboration affected by positive evidence.

5. Conclusion

The teacher of probability is bound to encounter some initial resistance that may originate from the deeply-rooted human quest for certainty. Every student must undergo a 'probabilistic revolution.' This involves understanding that questions of probability do not require prediction of immediate outcomes — that they refer to long-run relative frequencies. Teachers should be aware of the prevalence of misconceptions of chance and help students confront difficulties in assimilating the concept of statistical independence. Explication of the random mechanism that has generated the data given in a probability problem is one of the most important steps toward solving the problem. Despite the counterintuitive nature of some probabilistic results, many of probability theory's rules are compatible with everyday logic and are psychologically acceptable. We advocate capitalizing on such commonsensical inferences to establish students' confidence in their ability to reason probabilistically, and to help them see that reasoning as an extension of common-sense reasoning.

References

1. Bar-Hillel, M., and R. Falk (1982). Some teasers concerning conditional probabilities. *Cognition*, **11**, 109-122.
2. Clement, J., D. E. Brown, and A. Zietsman, (1989). Not all preconceptions are misconceptions: Finding 'anchoring conceptions' for grounding instruction on students' intuitions. *International Journal of Science Education*, **11** (special issue), 554-565.
3. Falk, R., R. Falk, and I. Levin, (1980). A potential for learning probability in young children. *Educational Studies in Mathematics*, **11**, 181-204.
4. Gigerenzer, G., Z. Swijtink, T. Porter, L. Daston, J. Beatty, and L. Krüger, (1989). *The empire of chance*. Cambridge: Cambridge University Press.
5. Gilovich, T., R. Vallone, and A. Tversky, (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, **17**, 295-314.
6. Kahneman, D., P. Slovic, and A. Tversky, (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
7. Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction*, **6**, 59-98.

8. Krüger, L., L. Daston, and M. Heidelberger, (Eds.). (1987). *The probabilistic revolution: Vol. 1. Ideas in history*. Cambridge, MA: MIT Press.
9. Krüger, L., G. Gigerenzer, and M. S. Morgan, (Eds.). (1987). *The probabilistic revolution: Vol. 2. Ideas in the sciences*. Cambridge, MA: MIT Press.
10. Phillips, L. D. (1973). *Bayesian statistics for social scientists*. London: Nelson.
11. Piaget, J., and B. Inhelder, (1975). *The origin of the idea of chance in children* (L. Leake, Jr., P. Burrell, and H. D. Fishbein, Trans.). New York: Norton (Original work published 1951).
12. Wagenaar W. A., and G. B. Keren, (1988). Chance and luck are not the same. *Journal of Behavioral Decision Making*, **1**, 65-75.

Ruma Falk is a psychologist at the Hebrew University of Jerusalem. She has taught probabilistic and statistical thinking for many years to students of psychology, social work and education. Her research deals with the judgment of randomness and coincidences, inference via conditional probabilities, and the development of the concepts of probability and infinity in children.

Cliff Konold is at the Scientific Reasoning Research Institute in the Hasbrouck Laboratory of the University of Massachusetts at Amherst. He is Principal Investigator of the NSF-funded project, "A Computer-Based Curriculum for Probability and Statistics". As part of this project, he has designed interactive software for use in introductory high school and college courses.