Cognition, 43 (1992) 197–223

A closer look at the probabilities of the notorious three prisoners*

Ruma Falk

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

Received April 15, 1991, final revision accepted December 5, 1991

Abstract

Falk, R., 1992. A closer look at the probabilities of the notorious three prisoners. Cognition, 43: 197–223.

The "problem of three prisoners", a counterintuitive teaser, is analyzed. It is representative of a class of probability puzzles where the correct solution depends on explication of underlying assumptions. Spontaneous beliefs concerning the problem and intuitive heuristics are reviewed. The psychological background of these beliefs is explored. Several attempts to find a simple criterion to predict whether and how the probability of the target event will change as a result of obtaining evidence are examined. However, despite the psychological appeal of these attempts, none proves to be valid in general. A necessary and sufficient condition for change in the probability of the target event, following observation of new data, is proposed. That criterion is an extension of the likelihood-ratio principle (which holds in the case of only two complementary alternatives) to any number of alternatives. Some didactic implications concerning the significance of the chance set-up and reliance on analogies are discussed.

Correspondence to: Ruma Falk, Department of Psychology, The Hebrew University, 91905 Jerusalem, Israel.

*This paper was written while the author was on sabbatical leave at the Department of Psychology, University of Massachusetts, Amherst. The study was supported in part by the Sturman Center for Human Development, the Hebrew University, Jerusalem, and by the National Science Foundation, Grant MDR-8954626 to Clifford Konold. The author wishes to thank Cliff Konold and Abigail Lipson for their thoughtful comments, and Raphael Falk for his continuous help and advice throughout all the stages of the study. Thanks are due to Gerd Gigerenzer and another anonymous reviewer for their helpful contributions.

Introduction

The problem of the three prisoners, a problem of apparently elementary structure, poses a serious challenge to common sense. Gardner (1961) and Mosteller (1965), both of whom have published many intriguing problems, report that the three prisoners bring the biggest flood of readers' letters.

> Tom, Dick, and Harry are awaiting execution while imprisoned in separate cells in some remote country. The monarch of that country arbitrarily decides to pardon one of the three. The decision who is the lucky one has been determined by a fair draw. He will be freed; but his name is not immediately announced, and the warden is forbidden to inform any of the prisoners of his fate. Dick argues that he already knows that at least one of Tom and Harry must be executed, thus convincing the compassionate warden that by naming one of them he will not be violating his instructions. The warden names Harry. Thereupon Dick cheers up, reasoning: "Before, my chances of a pardon were 1/3; now only Tom and myself are candidates for a pardon, and since we are both equally likely to receive it, my chance of being freed has increased to 1/2."
>
> Suppose, however, that the warden had named Tom. By the same reasoning, this piece of information would be equally encouraging for Dick. It looks like, whoever the warden names, Dick's chances are affected favorably. In fact, just *imagining* the potential exchange with the warden would have the same effect . . . Can all this be true? More than that, the warden need not actually exist. Just a thought experiment on Dick's part, involving a *hypothetical warden*, would raise Dick's probability of survival. What is true for Dick, however, is valid for Tom and Harry as well, so that each prisoner's probability of going free is raised to 1/2, thereby violating the requirement for the sum of probabilities of all elementary events in a discrete sample space.[1] (See Bar-Hillel & Falk, 1982, pp. 118–119; Falk, 1978, p. 68; Weintraub, 1988, pp. 169–170; Zabell, 1988a, pp. 334–335.)

In addition to being included in some of the most instructive collections of teasers in probability theory (Gardner, 1961; Mosteller, 1965; Székely, 1986), the three-prisoner problem has been the focus of discussion by several authors who approach it from various viewpoints. Beckenbach (1970), whose orientation is

[1]Despite the perplexing conclusions, the problem is not generally considered a paradox. It appears in Székely's (1986, pp. 68–69) book under the heading *Absurdities*, and is considered by the author to be a *fallacy*, that is, a nonsense conclusion obtained by erroneous reasoning that seems correct. Székely reserves the term paradox to describe a true though surprising theorem. Other authors, however, for example Zabell (1988a, p. 334), do label the problem of the three prisoners a paradox, apparently defining that term differently.

didactic, examines several possible chance set-ups leading to the facts given in the problem, and he represents them by combinatorial models. Diaconis and Zabell (1986) offer an extensive statistical analysis of the problem by suggesting three ways to model it using upper and lower probabilities. Weintraub (1988) approaches the same problem philosophically, highlighting its implications for confirmation and for the relationship between first- and second-order probabilities. Recently, Shimojo and Ichikawa (1989) have invented a new version of the problem that enabled them to distinguish among different solution strategies that subjects employ.

Many variations of that popular problem have been published. This fact may cause some confusion in communication. I will therefore refer in the subsequent discussion to the version given above and to a set of notations to be introduced.

The resolution of the apparent paradox entailed by the three-prisoner situation crucially depends on some assumptions that need to be explicated. Similarly, the most salient incorrect solutions are based on intuitive beliefs that are rarely questioned. It is of interest to look, one by one, at the beliefs invoked by this problem. The tenability of various possible beliefs and the psychological reasons why some of them are more likely to be embraced will be examined in this paper.

The three-prisoner problem is, in fact, representative of a class of counterintuitive probability puzzles. Different problems in that class may vary in their cover story, and some parameters may assume different numerical values. But, despite these surface variations, the basic mathematical structure of these problems is the same: we are presented with an uncertain target event of a given probability; additional information, supposedly concerning some other event in the sample space, is then provided, and we are asked about the revised probability of the target event.

In addition to trying to solve the problem, people tend to look for a simple and sensible criterion that will *always predict* whether and how the probability of the target event will change as a result of obtaining evidence. It turns out that none of the most intuitively appealing criteria is valid in general. After considering a few suggestions of that kind, I will describe a criterion, that is, a necessary and sufficient condition for change in the probability of the target event following observation of new data. Although that criterion is not primarily intuitive, it will be shown to make sense.

Resolving the problem

What is the error in reasoning that has led to such a paradoxical conclusion? And what is the actual probability of pardon for Dick, given the warden's statement that Harry is to be executed?

The expedient course in grappling with probability puzzles of this kind typically involves three major stages (Falk & Konold, in press).

(1) Uncover hidden assumptions, and check whether they are warranted (Bransford & Stein, 1984, pp. 85–88 and 94–97).
(2) Explicate the random process that has generated the data (Falk, 1983; Gardner, 1961; Zabell, 1988a).
(3) Apply Bayes' theorem.

The first and the second steps are obviously interwoven. They can be employed in parallel, since description of the underlying random mechanism is often based on some assumptions that ought to be made explicit. Let us apply these measures to the problem of the three prisoners.

(1) Dick's reasoning begins with the assignment of equal a priori probabilities to the three prisoners going free. This is, in fact, stated in the problem's text. The pardoned prisoner has been selected by a fair draw. It is important, however, to realize that the solution of the problem may depend on this assumption. As we shall see later, it is easy to overlook the dependence of the numerical answer on the specific conditions given in the problem. Other situations could be imagined in which the initial probabilities of pardon for Dick, Tom, and Harry are not equal (e.g., the probabilities of receiving the pardon could be inversely proportional to the severity of their respective offenses).

It is reasonably assumed, in addition, that the warden is truthful. The most important assumption, however, which directly entails the rise in Dick's assessment of his probability of survival, is that Dick and Tom remain equally likely to be the recipients of the pardon, once Harry has been ruled out via the information provided by the warden. That assumption is not "hidden"; it is, however, incorporated in Dick's reasoning as self-evident, and it is not questioned.

On closer examination it is clear that although the possible outcomes of the original fair draw which determines who will be pardoned are {Tom, Dick, Harry}, that sample space does not describe the contingencies when the warden's uncertain response is also considered. The potential statement of the warden is part and parcel of the chance situation. In order to be able to determine, in light of the new evidence, whether the posterior probability of Tom going free is equal to that of Dick going free, we have to carefully examine how that information has come about.

(2) The rules of the game, tacitly agreed upon by Dick and the warden, imply that the warden is not to mention Dick's name. This means that if Tom is pardoned the warden is certain to name Harry as the one about to die, and vice versa. If, however, Dick is the pardoned prisoner, the warden can name either Harry or Tom. How is the truthful warden supposed to behave when confronted with a choice? The text provides no clues. The readers must supply their own assumptions.

The first suggestion that comes to mind is to let chance decide, that is, to *flip a coin*. If it lands on Heads the warden names Harry, and if it lands on Tails he names Tom. In the absence of information about the warden's preferences, the simplest is to presume he is indifferent. This seems the fairest assumption. Now that the underlying chance procedure is clear, we can compute the posterior probability of Dick being freed.

(3) Let T, D, H denote the respective events that Tom, Dick, or Harry are pardoned. Given the equality of the prior probabilities of pardon for the three prisoners: $P(\mathrm{T}) = P(\mathrm{D}) = P(\mathrm{H}) = 1/3$, we now have to translate the terms of the understanding struck with the warden into symbols and assign numerical values to them. Let h be the event that the warden names Harry as the prisoner to be executed;[2] t will be the event that he names Tom. The following conditional probabilities describe the likelihoods of obtaining either testimony:

$$P(\mathrm{h}\,|\,\mathrm{T}) = 1\,; \quad P(\mathrm{h}\,|\,\mathrm{D}) = 1/2\,; \quad P(h\,|\,\mathrm{H}) = 0$$

$$P(\mathrm{t}\,|\,\mathrm{T}) = 0\,; \quad P(\mathrm{t}\,|\,\mathrm{D}) = 1/2\,; \quad P(\mathrm{t}\,|\,\mathrm{H}) = 1$$

Once we are informed that the warden has named Harry, we are interested in determining $P(\mathrm{D}\,|\,\mathrm{h})$: the probability that Dick has been pardoned given that information. By Bayes' theorem:

$$P(\mathrm{D}\,|\,\mathrm{h}) = \frac{P(\mathrm{h}\,|\,\mathrm{D})P(\mathrm{D})}{P(\mathrm{h}\,|\,\mathrm{T})P(\mathrm{T}) + P(\mathrm{h}\,|\,\mathrm{D})P(\mathrm{D}) + P(\mathrm{h}\,|\,\mathrm{H})P(\mathrm{H})} \tag{1}$$

$$= \frac{(1/2)(1/3)}{1(1/3) + (1/2)(1/3) + 0(1/3)} = \frac{1}{3}$$

Dick's chances of receiving the pardon are thus unchanged from what they were before learning about Harry's fate. Tom's chances, however, are no longer equal to Dick's. They are now 2/3.

The seemingly innocuous assumption that Tom and Dick remain equally likely to be pardoned, subsequent to learning from the warden that Harry will be executed, has thus proved wrong. That assumption, which has speciously been incorporated in the argument, is evidently deceptive. It is the heart of the ensuing paradoxical conclusions. Although it is tempting to persist in postulating uniformity, this assumption should be entertained with a fair amount of suspicion. Indeed, it is often the case in Bayesian analysis that equally likely events assume

[2]A more elegant notation for that evidence would be $\bar{\mathrm{h}}$, since H denotes the event that Harry is pardoned and consequently $\bar{\mathrm{H}}$ means that he will be executed. However, the notation h, signifying the warden's statement that Harry will be executed, will be used for the sake of simplicity.

unequal probabilities when revised in the light of new observations (Bar-Hillel & Falk, 1982).

The lack of change in $P(D)$ does not imply, however, that the warden's statement is useless information. Gardner (1961) vividly demonstrates this point by imagining that Dick surreptitiously communicates with Tom, in an adjacent cell, by tapping in code on a water pipe. Dick transmits to Tom an exact account of his exchange with the warden. On hearing this news, Tom should be overjoyed. As we have seen, his chances of survival now rise, not to 1/2, as Dick has figured, but to 2/3 (see also Zabell, 1988a).

Our revised probabilities, conditioned on the warden's testimony, are now $P(T \mid h) = 2/3$; $P(D \mid h) = 1/3$; $P(H \mid h) = 0$. Do these results make sense? In the next section we shall have a look at some prevalent intuitions and try to better understand why the three-prisoner problem and its solution are sometimes counterintuitive. I will also examine intuitions that are compatible with the correct solution and try to find out whether they reflect generally valid principles, or whether they happen to work due to fortuitous circumstances.

Primary intuitive beliefs

The beliefs that solvers bring to this problem could refer to two levels. People could (1) presume some underlying procedure that has resulted in the warden's statement, or (2) apply some intuitive assumptions directly to the question of Dick's posterior probability of survival given the warden's testimony. In fact, it is rare that naive solvers devote any thought to the former. They usually perceive no need to specify procedural assumptions or to take into account the probabilities associated with the warden's choice. Beliefs relating to the latter question are common, however, and most of them are erroneous.

The two most prevalent intuitive beliefs are actually included in the problem's text. The first, as mentioned earlier, is the equiprobability of the remaining alternatives, that is, the *uniformity* belief. The second, mentioned in passing, is embedded in Dick's attempt to induce the warden to tell him the name of one of the other prisoners who will be executed. Dick argues that he already knows that at least one of Tom or Harry is to be executed. Therefore, he maintains, he will receive no information about his own fate by getting one of these two names. If nothing new has been learned, no change in the probabilities should ensue. This is the "*no-news, no-change*" (or "I've known it all along") belief.

The uniformity belief is seldom questioned. It appears to be second nature for many naive, as well as statistically educated solvers. As for the "no-news" argument, it has convinced not only the warden, but most of the readers of the three-prisoner problem as well. It leads to a probability of pardon of 1/3 for Dick

(the correct answer for this particular problem), whereas the uniformity assumption entails an answer of 1/2. The same two intuitions repeatedly come up in connection with the three-prisoner problem and with analogous problems. Shimojo and Ichikawa (1989) regard them as "subjective theorems", and they label them "Number of cases theorem" and "Irrelevant, therefore invariant theorem", respectively. They report that a vast majority of their subjects endorse one or the other of these beliefs and base their solution thereupon. This is true even when the cues to these two assumptions are removed from the problem's text. Having myself posed that problem to many students and colleagues, I can readily corroborate the ubiquity of the two beliefs.

These two beliefs, however, clash with each other, so that whoever endorses both of them is bound to be perplexed. Interestingly, many respondents manage to avoid conflict by strictly clinging to one of the arguments, plainly ignoring the other one (thereby deeming it incorrect by default). When counting votes, the uniformity argument seems to "win", as illustrated by a recent public discussion concerning an analogous problem (Morgan, Chaganty, Dahiya, & Doviak, 1991; Saunders, 1990).

In a famous TV game show, "Let's Make a Deal" (see Selvin, 1975a, 1975b), the contestant is given a choice of three closed doors. Behind one of the doors is an attractive prize (e.g., a car); behind the other two are gag prizes (e.g., goats). Suppose the contestant picks door no. 1. It remains closed and the host, Monty, who knows what's behind the doors, opens one of the other doors, say no. 3, to reveal a goat. He then gives the contestant the option of switching to no. 2. Should the contestant stick or switch? (See Shaughnessy & Dick, 1991.)

The similarity of "Monty's dilemma" to the problem of the three prisoners is evident. In fact, the mathematical structure of the two problems is identical. Monty's dilemma along with the solution was published in the "Ask Marilyn" column of *Parade* magazine. Marilyn correctly advocated a switch since door no. 1 has a 1/3 chance of winning whereas no. 2 has a 2/3 chance. Marilyn reports receiving thousands of letters from the readers, including many from universities and research institutes, about 90% of them insisting that she was wrong. Many of the published letters smugly reprimanded her for being in error (*Parade* magazine, December 2, 1990, p. 25 and February 17, 1991, p. 12).

Luckily, the truth of mathematical propositions is not decided by preponderance of votes. Incidentally, however, we have gained from these statistics an important psychological lesson about the pervasiveness of the uniformity assumption, at least among the group of people who chose to write in. Although that group might constitute a biased sample of the population of readers, it stands to reason to assume that the bias cannot be so extreme that it would reverse the count of opinions from a minority to an overwhelming majority favoring the equiprobability of the remaining alternatives.

The uniformity assumption

Marilyn's attackers all believe in the equiprobability of the two remaining possibilities. That is why they think there is no justified reason to switch. Here are two examples:

(1) If one door is shown to be a loser, that information changes the probability to 1/2. As a professional mathematician, I'm very concerned with the general public's lack of mathematical skills. Please help by confessing your error and, in the future, being more careful.
(2) You blew it, and you blew it big! I'll explain: after the host reveals a goat, you now have a one-in-two chance of being correct. Whether you change your answer or not, the odds are the same. There is enough mathematical illiteracy in this country, and we don't need the world's highest IQ [referring to Marilyn, the column's editor] propagating more. Shame!

Additional documentation for the prominence of the uniformity belief can be found in Bar-Hillel and Falk's (1982, p. 119) report where a majority of the subjects (66%) responded erroneously in accord with that belief to an analogous probability problem (the three-card problem). Equiprobability prevailed though it did not compete with a "no-news" argument in that case. Results pertaining directly to these two rival beliefs, in connection with the three-prisoner problem, are obtained by reanalyzing Shimojo and Ichikawa's (1989) data. Their three experiments employed 125 students as subjects (see their Table 2), divided into six groups. Throughout the three experiments, 77 subjects responded twice to the traditional form of the three-prisoner problem (the form that does not introduce unequal priors) and 48 of them responded once. Altogether, 202 responses had been collected (variations in order of presentation and in details of the problem's formulation are ignored since they were found to have no effect). One hundred and twenty of the pooled 202 answers (i.e., 59.4%) were 1/2, endorsing the equiprobability belief, whereas 75 (i.e., 37.1%) were 1/3, in accord with the "no-news" belief. Only 3.5% of the answers were based on neither of these beliefs.

Those believing in posterior uniformity are not only numerous, there is also a special quality to their conviction: they are highly confident. Their belief can be considered a *primary intuition* since it is marked by some of the most distinctive features of such a cognition. Fischbein (1987, pp. 200–201) characterizes intuition in terms of *self-evidence and immediacy*. Indeed, people rarely display any shred of doubt when they instantaneously rely on the uniformity assumption, as if there is *intrinsic certainty* to that belief. Intuitive beliefs (according to Fischbein) exert a *coercive* effect on the individual's reasoning and choice of strategy. Intuition is also characterized by *perseverance* in being resistant to alternative arguments.

One can find a plethora of attempts to define "intuition" in the literature (e.g., Delbrück, 1986, p. 277; Fischbein, 1987, pp. 13–14, 43–56, 200–202; Kahneman & Tversky, 1982, p. 494). Yet, the concept seems to evade a rigorous definition.

According to Delbrück "there is a vast philosophical literature concerning this particular word, which suggests that its meaning is vague enough to cover a multitude of sins" (1986, p. 277). "Intuitive belief" is used here in an intuitive way, in a sense closest to that described above by Fischbein (1987). An intuitive belief turns into a *heuristic* when it is employed as a rule of thumb to solve a problem.

The primacy of the uniformity intuition can be traced historically to early stages in the development of probability theory. It is commonly supposed that the assumption of equally likely cases originated with Laplace around the end of the eighteenth century, but in fact it was commonplace as early as a century before that[3] (Hacking, 1975, chapter 14). Laplace defined probability as the ratio of favorable cases to the total number of equally possible cases. The heart of the mandate given by Laplace to reduce the computation of probability to that of relative frequency was the conception that cases of which we are ignorant in the same way are equiprobable.[4] The interpretation of probability based on that assumption was in full vigor a century after Laplace and is still much with us.

A recent example is comfortably provided by Shimojo and Ichikawa (1989). They justify their assigning equal prior probabilities of survival to their three prisoners by relying on the lack of any information indicating otherwise (there was no fair draw in their story), namely, on the "principle of insufficient reason" (p. 2).

In a study on college students' conceptions of randomness, Konold et al. (in press) found that undergraduate students of psychology typically agreed that blindly picking a white marble from a box that contains 10 black and 10 white marbles is a random event. Only 70% of them, however, maintained that picking a white marble from a box that contains 10 black and 20 white marbles is a random event. It is noteworthy that even though drawing marbles from urns is prototypically considered a random process, some students regard it as nonrandom because they equate chance with uniformity. They consider a phenomenon random only when all its outcomes are equally likely.

By the same token, Zabell (1988b) reports that in the early days of the doctrine of chances some thought the notion of chance only applicable to partitions of the space of possible outcomes into equiprobable alternatives. Based on official London statistics of consecutive 82 years in which male births had

[3]Leibniz is known to have relied on uniformity in 1678.

[4]D'Alambert (1717–1783) was once asked the following question: what is the probability of obtaining at least one head when a fair coin is tossed twice? D'Alambert's answer is reported (Székely, 1986, p. 3) to be 2/3 (the correct answer is 3/4). He reasoned that if heads appears first then the game is over and there is no need for a second trial. His sample space consisted of three outcomes: H, TH, TT, and he assumed erroneously that these outcomes were equally likely (Glickman, 1986). Similar mistakes are made quite frequently even nowadays.

exceeded female births in each single year, Arbuthnot rejected, in 1710, the hypothesis of equilikelihood, making in effect the earliest known statistical test of significance. But Arbuthnot did not conclude that male and female births possessed unequal probabilities. Instead, he rejected outright the possibility that sex was due to chance, concluding that the excess of males was due to the intervention of divine providence (see also Hacking, 1990, p. 21: "This could not result from chance (i.e., equal chances)"). The belief that chance is only operative when outcomes are equilikely is echoed in Hume, as quoted by Zabell (1988b): "an entire indifference is essential to chance, no one chance can possibly be superior to another" (p. 171).

Gigerenzer et al. (1989) relate that both Bayes and Laplace made the problematic assumption that in the absence of any information to the contrary, we may convert ignorance into a uniform distribution of probabilities. However, while Bayes appeared to have been troubled by the assumption (and made elaborate attempts to justify it), Laplace was nonchalant in presuming uniformity, offering no justification whatsoever, much like present-day solvers of the three-prisoner (and Monty's) problem. One should note, however, that the historical assumption of uniformity referred to the distribution of prior probabilities, whereas the taken-for-granted belief in the case of the three prisoners assigns equal posterior probabilities to all possible cases.

The mere supremacy of the equiprobability presumption throughout history does not explain the psychological proclivity toward that assumption. Hacking (1975) regards the phenomenon as puzzling: "Here is an historical problem. How could so monstrous a definition have been so viable? . . . why so dubious a concept as equipossibility should have had such a successful career in well over two centuries of lively theorizing" (p. 122). Hacking's explanation is based (first of all) on the essential duality of probability, which is both epistemic and aleatory. Aleatory probabilities have to do with the objective, physical chance set-up under consideration (such as coins, or mortal humans in risk). Epistemic probabilities concern our state of knowledge (see chapter 14 for an exposition of Hacking's analysis).

Though a fair draw, of a seemingly physical nature, has been introduced into the three-prisoner story in order to stress the a priori equality of the probabilities of pardon, the puzzling consensus about the equality of the posterior probabilities has to be understood in the framework of the epistemic interpretation of probability. To assume uniformity, when we think we know nothing else but the list of possible outcomes, seems so natural that just describing the phenomenon seems sufficient to explain it. Dividing the uncertainty equally among the available possibilities can perhaps be understood in terms of the sensibility of historical notions such as the canons of "insufficient reason", "equal ignorance", and the "principle of indifference". All these might reflect a basic preference for *symmetry* and *fairness*.

While it is debatable whether total ignorance should be translated to equal probabilities, people are surely in error in overlooking the differential impact of the warden's statement on the chances of the remaining possibilities. The evidence given by the warden, if properly assimilated, should change our epistemic state of ignorance. We should now distribute the probabilities of pardon unequally between Dick and Tom.

The "no-news" assumption

Judging by preponderance, the uniformity assumption is ranked topmost, despite a marked difficulty to defend it if asked. If, on the other hand, we grade assumptions by their persuasiveness, no-news has no rival. At face value, there is something compelling about the idea that when we receive a piece of information we have known before, it should not alter our assessment of the situation. Obviously, Dick knows from the start that the warden can truthfully name either Tom or Harry. That being the case, how could hearing either of these names change Dick's view? The argument that the mere utterance of one of these names makes no difference for Dick is indeed powerful. More than that, formal calculation has affirmed the solution resulting from the no-news assumption (i.e., $P(\mathrm{D} \mid \mathrm{h}) = 1/3$).

Marilyn relies on the same argument when trying to convince her stubborn correspondents that the uniformity assumption cannot be applied to the situation subsequent to opening door no. 3:

> The winning odds of 1/3 on the first choice can't go up to 1/2 just because the host opens a losing door. To illustrate this, let's say we play a shell game. You look away, and I put a pea under one of three shells. Then I ask you to put your finger on a shell. The odds that your choice contains a pea are 1/3, agreed? Then I simply lift up an empty shell from the remaining two. *As I can (and will) do this regardless of what you've chosen* [italics added], we've learned nothing to allow us to revise the odds on the shell under your finger (*Parade* magazine, December 2, 1990, p. 25).

Truly, Marilyn can always lift an empty shell, but she did not specify which of the two empty shells she would lift if the pea is under the player's finger. Suppose she is biased, because of some idiosyncratic reason of her own, toward lifting one of the two shells (and you know about that bias). She'll always lift this one when she can. The only time she would lift the other one would be when the pea was under her preferred shell. In that case, it is still true that she can (and will) always reveal an empty shell. Yet, it is not true that having lifted the shell of her preference, "we've learned nothing to allow us to revise the odds on the shell under your finger". Considering that act, the probability that the pea is hidden under your finger rises from 1/3 to 1/2. If, however, Marilyn lifts her unfavored shell, you know that the pea is under her preferred shell, and the probability that it is under your finger drops to zero.

Getting back to the language of the three prisoners, we can now derive these results formally. Marilyn's assumed bias will be translated to assuming the warden is biased toward naming Harry, that is, $P(\mathrm{h}\,|\,\mathrm{D}) = 1$, and consequently $P(\mathrm{t}\,|\,\mathrm{D}) = 0$. All the rest is as before. By Bayes' formula (1), having the warden name Harry will now result in

$$P(\mathrm{D}\,|\,\mathrm{h}) = \frac{1(1/3)}{1(1/3) + 1(1/3) + 0(1/3)} = \frac{1}{2}$$

and, as can easily be seen, naming Tom results in $P(\mathrm{D}\,|\,\mathrm{t}) = 0$.

Another variation of the problem – leaving the warden indifferent as to naming Tom or Harry, but changing the priors – is offered by Shimojo and Ichikawa (1989, p. 5). The priors in their version, retaining our notations, are $P(\mathrm{T}) = 1/2$, $P(\mathrm{D}) = 1/4$, and $P(\mathrm{H}) = 1/4$. Otherwise everything is as before. Substituting these values in (1), we obtain:

$$P(\mathrm{D}\,|\,\mathrm{h}) = \frac{(1/2)(1/4)}{1(1/2) + (1/2)(1/4) + 0(1/4)} = \frac{1}{5}$$

Even though we have known all along that the warden can name either Tom or Harry, his naming Harry does affect Dick's chances in this case. The probability that Dick is pardoned has surprisingly dropped from 1/4 to 1/5, thereby refuting the no-news notion. By the same token, if the warden names Tom (rather than Harry), the Bayesian calculation yields $P(\mathrm{D}\,|\,\mathrm{t}) = 1/3$, raising Dick's survival chances from 1/4 to 1/3. In both cases, however, the warden's statement is "news" for Dick. The refutation of the no-news argument has been the most important "news" for me in studying these problems.

In the original version of the problem, the warden can always truthfully name at least one of Tom and Harry, *and* the probability of pardon for Dick does not change. Yet, as we have seen, it is *not because* of the former that the latter is true. The invariance of Dick's chances of pardon can conceivably be traced to some particular combination of numerical parameters. The nature of that combination is of interest, and it will be explored later.

The intuitive appeal of the no-news reasoning is evidently irresistible. Furthermore, that belief turns into a conviction when it happens to be verified by the results of the formal calculation. This might be what Mosteller (1965) means when, after assuming the warden is unbiased and arithmetically deriving an unchanged target probability, he remarks: "and mathematics comes round to common sense after all" (p. 29).

Similarly, Gardner (1961) presents a lucid elicitation of the correct solution, based on equal priors and on assuming an indifferent warden; nonetheless, he explains: "Regardless of who is pardoned, the warden can give A the name of a man, other than A, who will die. The warden's statement *therefore* [italics added]

has no influence on A's survival chances; they continue to be 1/3" (p. 229). Both parts of that sentence are correct, just the adverb "therefore", used here with conjunctive force, is inapt[5] (see also Gardner, 1992). Evidently, none of us, even the best of experts, is immune to that fallacy.

Secondary intuitive beliefs

To this point I have been treating primary intuitions, which seem to be formed spontaneously in people's minds, irrespective of whether they are alluded to in the problem's story or not. *Secondary intuitions* also often turn up in dealing with the three-prisoner problem. These are semi-intuitive heuristics that are arrived at through some deliberations, and which seem plausible once they are formulated. Secondary intuitive beliefs are acquired partly as a result of instructional intervention. But they are also compatible with our anticipations, and they "make sense" (Fischbein, 1987).

The constant-ratio belief

The *constant-ratio belief* is the assumption that when some alternatives are ruled out, the ratio of the probabilities of the remaining alternatives should be the same as the ratio of their prior probabilities. This belief is obviously not valid since it leads to the answer 1/2, whereas the correct answer (assuming an unbiased warden) is 1/3.

In the original formulation all three prisoners are a priori equally likely to be pardoned. Thus, those who wish to leave the ratio of Dick's to Tom's chances unchanged, after eliminating Harry, will end up assigning equal posterior probabilities to the two. The answer 1/2, however, does not distinguish between users of the uniformity versus the constant-ratio heuristic in the classical version of the problem.

Versions starting with unequal prior probabilities permit discrimination between the two heuristics. In Shimojo and Ichikawa's (1989, p. 5) modified version – where $P(\mathrm{T}) = 1/2$, $P(\mathrm{D}) = 1/4$, and $P(\mathrm{H}) = 1/4$ – the uniformity assumption leads to the answer 1/2, while the constant-ratio belief results in 1/3 (no-news entails 1/4, and the correct answer, to be recalled, is 1/5). Indeed, Shimojo and Ichikawa report (p. 11) that about half those given this version gave

[5]To be fair, Gardner (1961, p. 230) resorts to analogous card problems in order to better explain his solution. In that explanation he does mention the dependence of the solution on two different methods of deciding which cards to turn face up (which prisoner to name). Nevertheless, he maintains the causal connection between being always able to turn over a black card and obtaining no information of value in betting on the target event.

1/3 as their answer, and most of them clearly endorsed the constant-ratio belief. (The other half gave the answer 1/4, apparently favoring the no-news view.)

What is the basis of the belief that the ratio of the probabilities should remain constant? Several possible reasons seem to contribute to making that assumption available to subjects. In a way, embracing the constant-ratio belief is overdetermined.

First, the constant-ratio belief may be a generalized form of the uniformity assumption, extended to the case of unequal prior probabilities. The reader may be disturbed by the ostensible contradiction between "uniformity" and "unequal probabilities". All the same, one could speculate that the very quest for fairness, which brings about the uniformity assumption in the absence of any information to the contrary, leaves the ratio of uncertainties unchanged when unequal priors are the only available information. Being aware of the initially nonuniform distribution, and of the need to adjust the probabilities to add up to one, following the elimination of one alternative, the problem solver distributes the remaining uncertainty "impartially" among the existing alternatives, thus preserving the same ratio.

Second, those who believe that the warden's evidence is not informative and should therefore induce no change in Dick's views, may nevertheless adopt the constant-ratio strategy in order to adjust the sum of the probabilities. They apply the no-news no-change notion to the *ratio* of the remaining probabilities, whose sum should be one. On the whole, the constant-ratio belief can be viewed as a refinement of both the uniformity and the no-news no-change assumptions.

Finally, children are usually exposed, quite early in the course of their scholastic track, to the concept of proportion. Four numbers, a, b, c, d, are in proportion when the ratio of the first pair equals the ratio of the second pair. This is denoted by $a:b = c:d$. Students are drilled in the so-called "rule of three", namely, the method of finding the fourth term of a proportion given three terms. In the course of our scholastic education we repeatedly encounter the need for proportional computation, whether in solving arithmetic word problems of different contexts or in geometry and physics. One could speculate that this is an overlearned principle that is too readily available. Hence the decision to apply the same rule also to the three-prisoner situation, where two of the four numbers are given and the other two should add to one.

Historically, the birth of probability theory is attributed to the Pascal–Fermat correspondence concerning two famous problems. These problems were posed by the Chevalier de Méré, and were solved independently by the two mathematicians in 1654. The problems proved difficult because of the failure of the traditional rule of three to provide an adequate solution. The proportional principle was so well established at the seventeenth century that these problems were considered "paradoxes" (Freudenthal, 1970; Glickman, 1986, 1989; Székely, 1986).

Here is one version of the story involving the first problem:

> De Méré knew that it was advantageous to bet on the occurrence of at least one six in a series of four tosses of a die – maybe this was an old experience. He argued it must be as advantageous to bet on the occurrence of at least one double-six in a 24 toss series with a pair of dice. As Fortune disappointed him, he complained to his friend Pascal about preposterous mathematics which had deceived him (Freudenthal, 1970, p. 151).

De Méré was confused by the fact that his observed results did not conform to the taken-for-granted proportion $4:6 = 24:36$. He expected 4 tosses of one die, with 6 possible outcomes, to be probabilistically equivalent to 24 tosses of two dice, with 36 possible outcomes. In fact, the probability of at least one six in four tosses is $1 - (5/6)^4 = 0.518$, somewhat more than one half, while that of at least one double-six in 24 double-tosses is $1 - (35/36)^{24} = 0.491$, somewhat less than one half. Glickman (1989) advocates introducing this historical problem to the classroom to caution the students against uncritically employing simple proportional ideas in probabilistic reasoning.

The second problem, known as the "Division Paradox" (Székely, 1986, pp. 9–11), concerns two players who play a fair game.[6] It is agreed that the first to win 6 rounds takes the stakes. The game gets interrupted after player A has won 5 rounds, and player B 3 rounds. How should the stakes be divided fairly?

The problem provoked dispute and confusion. Contradictory answers created the legend of a paradox. Pascal and Fermat considered it a problem of probabilities. They ruled, as arbiters, that the fair division would be according to the ratio of the probability of player A winning to that of player B (were they to continue playing). That ratio is 7:1.

The exact derivation of these probabilities is not as important as the prominence of solutions involving simple proportional reasoning. One of these was to divide the stakes in the ratio of rounds won, 5:3; others were in favor of (6–3):(6–5). Common to all the suggestions was an attempt to preserve some simple proportion given in the data.[7]

[6]The Division Paradox was first published in the fifteenth century by Fra Luca Paccioli, but there are indications that it was known much earlier. Paccioli himself did not even realize its connection with probability theory, for he considered it simply a problem in proportions. (Paccioli was the author of "De Divina Proportione", illustrated by his close friend Leonardo da Vinci, which was published in Venice in 1509.)

[7]A similar "belief" in constant-ratio was manifested in a letter, written in 1693 by Samuel Pepys to Isaac Newton, inquiring whether the probabilities of the following three events are equal: obtaining at least one ace when throwing six dice; obtaining at least two aces when throwing twelve dice; obtaining at least three aces when throwing eighteen dice (Glickman, 1986). The idea behind that query is remarkably similar to that of de Méré's first problem, namely, should not the probabilities of the three events be equal since $1:6 = 2:12 = 3:18$? (The fact of the matter is they are not. The first is most probable.)

Freudenthal (1970) comments on these examples:

> Chevalier de Méré certainly was an educated man, no doubt he had learned mathematics, yet . . . he applied the mathematics he knew, the mathematics of what in my infancy was called the rule of three . . . De Méré is not so much a historical figure. He is a paradigm, the ancestor of a prolific offspring, of all the poor de Mérés of our days . . . From the oldest times the rule of three or in modern terminology, the linear function has been an important mathematical tool of explaining and mastering phenomena in physics, chemistry, astronomy, economics, and in any field of human activity. It is a cheap procedure . . . using it dispenses with rethinking a situation, and such a dispensation is gladly accepted (pp. 152–153).

No wonder, then, that some people are so prone to embrace the constant-ratio assumption when faced with the modified version of the problem of the three prisoners where the prior probabilities are unequal.

The symmetry heuristic

Given that the apparent no-news situation does not guarantee no-change, and the initial ratio between Dick's and Tom's probabilities does not predict their posterior ratio, the natural question is whether there exists any generally valid cue that can predict when Dick's survival probability will change with the news about Harry's impending execution. Shimojo and Ichikawa (1989) inserted the likelihoods $P(h \mid T) = 1$, $P(h \mid D) = 1/2$, and $P(h \mid H) = 0$ (using our notation) in Bayes' formula (1) for computing Dick's posterior survival probability, and then solved for the mathematical condition that would render $P(D \mid h)$ equal to $P(D)$ (namely, no change in Dick's survival probability). They obtained $P(T) = P(H)$ as their condition (see Shimojo & Ichikawa, 1989, p. 20 and Appendix 1). Let us call this condition the *symmetry condition*.

Note that prior to telling us about the warden's testimony concerning Harry's fate, the story does not discriminate in any way between Harry and Tom. Only Dick has been singled out by addressing the warden. In particular, since $P(H) = P(T)$, it stands to reason that "by symmetry" naming *any* one of the two candidates (Harry or Tom) will have the same effect on Dick's prospects. Since it is impossible that the occurrence of each of two complementary events will increase (or decrease) Dick's probability, it follows that naming either of the two will no change $P(D)$.

The symmetry condition holds in the original version of the problem, where $P(T) = P(H) = 1/3$, and so does the invariance of Dick's probability for pardon. It should be pointed out, however, that the equality $P(T) = P(H)$ is not by itself a sufficient condition for securing no change in the target probability (nor is it a necessary condition, for that matter). The derivation of this condition has depended crucially on the specific triplet of conditional probabilities (likelihoods). Shimojo and Ichikawa (1989) are right in claiming that the symmetry condition is

sufficient for invariance of Dick's probability *provided* we assume the warden is indifferent, when Dick is to be freed, about naming Tom or Harry. If we drop that assumption, the symmetry condition is no longer sufficient.

Suppose the situation is changed only in that the warden will always answer "Harry" when he has a choice between Tom and Harry. Because the warden does now discriminate between Tom and Harry, $P(h|D)$ changes from 1/2 to 1. Replacing the old value by the new one in formula (1) yields $P(D|h) = 1/2$. Dick's survival probability has thus changed (from 1/3 to 1/2) in spite of the equality of $P(T)$ and $P(H)$, thereby refuting the generality of the symmetry heuristic.

If, however, we keep the initial equal probabilities and insist on assuming the warden is unbiased (i.e., $P(h|D) = 1/2$), then *everything* is *symmetrical* with respect to Tom and Harry, and naming either of them cannot affect Dick's chances. The argument is now impeccable. That set of conditions, which can be labeled the *complete symmetry*, is surely sufficient (though not necessary) for obtaining no change in Dick's survival probability. All the requirements of complete symmetry are satisfied in the classic three-prisoner story supplemented by the assumption of an unbiased warden. If it is to this constellation that Mosteller (1965) refers in saying, after obtaining the answer 1/3, "and mathematics comes round to common sense after all", then he is perfectly right.

The likelihood-ratio heuristic

Consider two opaque urns, W and B, each containing 100 beads. Urn W comprises 99 white beads and one black, and urn B 99 blacks and one white. One of the urns is randomly presented to you, so that a priori $P(W) = P(B)$. You blindly draw a bead from the urn, and it turns out white, denoted w. What is the probability that the selected urn is W? Formally, what is $P(W|w)$?

It makes sense to expect that urn W will become more probable than B as a result of observing w. That is so because the likelihood of observing w is greater when drawing from W than when drawing from B. Formally, transforming Bayes' formula so as to obtain the ratio between the posterior probabilities, we have

$$P(W|w)/P(B|w) = [P(w|W)/P(w|B)][P(W)/P(B)]$$

The answer is $P(W|w) = 0.99$. In this particular example, the ratio of the posterior probabilities equals the likelihood ratio, because $P(W) = P(B)$. In general, what determines which probability will increase (at the expense of the other one) is the ratio between the likelihoods of the observation, given the two alternatives.

When only two complementary situations, A and $\bar{A}$, are possible, it is always

true that observation of any datum, denoted d, satisfying $P(d \mid A) > P(d \mid \bar{A})$, will result in increase in A's probability, that is, in $P(A \mid d) > P(A)$. The same is true if we replace ">" systematically by "<" or by an equality sign (Phillips, 1973). It is therefore tempting to jump to the conclusion that a similar criterion may also be applied when three competing alternatives have been reduced to two after ruling out one via the observed datum.

The *likelihood-ratio heuristic*, in the case of the three prisoners, is the strategy of examining the ratio of $P(h \mid D)$ to $P(h \mid T)$; if it exceeds 1, we predict that Dick's survival probability will increase; if it is smaller than 1, we predict a decrease; and if it equals 1, we predict no change. For those who are familiar with Bayesian deliberations, this heuristic no doubt seems reasonable.

Unfortunately, the three prisoners manage to defy even this intuition. In the original version we have $P(h \mid D) = 1/2$ and $P(h \mid T) = 1$. However, despite the inequality $[P(h \mid D)/P(h \mid T)] < 1$, the probability of D stays unchanged[8] after obtaining the evidence h.

To this point, a common-sensical criterion to predict whether (and how) Dick's survival probability will change as a result of evidence has evaded us. In the following section, a valid criterion will be suggested.

The weighted-average criterion

A straightforward method to find a necessary and sufficient condition for invariance of Dick's survival probability, following the observation h, is to algebraically extract such a condition from Bayes' formula. It is easy to see that $P(D \mid h) = P(D)$ whenever the factor multiplying $P(D)$, on the right side of (1), equals 1, that is, whenever

$$P(h \mid D) = P(h \mid T)P(T) + P(h \mid D)P(D) + P(h \mid H)P(H) \tag{2}$$

If we now isolate $P(h \mid D)$, while replacing $1 - P(D)$ by $P(T) + P(H)$, we get

$$P(h \mid D) = \frac{P(h \mid T)P(T) + P(h \mid H)P(H)}{P(T) + P(H)} \tag{3}$$

[8]Even though there was no change in Dick's survival probability, the ratio of Dick's to Tom's probability did decrease, relative to the ratio of their priors, as a result of the evidence h (because Tom's survival probability had increased from 1/3 to 2/3). There is thus a core of truth in the intuition that the ratio $P(h \mid D) : P(h \mid T)$ should play an important role in updating our beliefs. The likelihood ratio, however, can only predict the direction of change in the *ratio* of Dick's to Tom's probability. It cannot, by itself, determine the revision of Dick's probability. The impact of the likelihood ratio is expressed by the equality $P(D \mid h)/P(T \mid h) = [P(h \mid D)/P(h \mid T)][P(D)/P(T)]$. Note that the ratio of the posterior probabilities is determined by the likelihood ratio as well as by the ratio of the priors. Thus, the "likelihood-ratio heuristic" and the "constant-ratio belief" each captures part of the truth, although neither is generally valid on its own.

This equality, labeled *the weighted-average criterion*, is a necessary and sufficient condition for $P(D \mid h) = P(D)$. This is the sought-after general criterion for invariance.

Despite the apparent complexity of this criterion, its meaning is quite simple. It requires that the likelihood of the evidence (i.e., h), given the target event (i.e., D), be equal to the weighted mean of the other (two) likelihoods (of h). The weights are the probabilities of the respective conditioning events. Furthermore, the reader can verify that when $P(h \mid D)$ is smaller than that weighted average, the impact of the observation h will be to decrease Dick's survival probability; and when $P(h \mid D)$ exceeds that weighted average, Dick's probability will rise.

Once the weighted-average criterion is derived, it becomes "intuitive". It is the correct extension of what is true in the case of two complementary possibilities, to the case of three (mutually exclusive and exhaustive) or to *any* other *number* of alternatives. When only two situations are possible, the weighted-average criterion is reduced to ruling that the impact of observation h is to increase the probability of that alternative conditioned on which h is more likely. When more than two alternatives compete, we pit $P(h \mid D)$ against a composite measure of the other likelihoods.

Looking back at the derivation of the weighted-average criterion (3), we can see that the condition for invariance, that is, the requirement that $P(D)$ in (1) be multiplied by 1, is equivalent to the condition that $P(h \mid D)$ would equal the weighted average of *all* the likelihoods of h, see (2), which is equal to $P(h)$. Viewing the criterion that way may help one to see the logic of it. It means that obtaining evidence h, whose likelihood under D is greater than its total probability, will raise D's probability.

The implicit assumption that the warden is unbiased, in conjunction with the terms of the warden's task, had led to the three likelihoods: $P(h \mid T) = 1$; $P(h \mid D) = 1/2$; $P(h \mid H) = 0$. Since $P(h \mid D)$ is equal to the arithmetic mean of the other two likelihoods, it follows that when the weights $P(T)$ and $P(H)$ are equal, $P(h \mid D)$ will equal the weighted average of the other likelihoods, resulting in no change in Dick's probability. That situation is identical to what has been described above as *the complete symmetry*. Indeed, when everything concerning Tom and Harry is symmetrical, it follows logically that Dick's chances are not affected by naming either. We see now, however, that complete symmetry is not the only case of no change in the target probability. The weighted-average condition for invariance holds for infinitely many different combinations of values of likelihoods and weights.

Another example we have considered was that of the biased warden who will name Harry when he can. This changes only $P(h \mid D)$ from 1/2 to 1. The weighted average of the other likelihoods remains 1/2, which is less than $P(h \mid D)$. Our criterion predicts a rise in Dick's survival probability which was initially 1/3. This is born out by the result $P(D \mid h) = 1/2$ for that case. Different degrees of bias for

or against naming Harry can be expressed by the values that $P(h \mid D)$ assumes, and the direction of change in Dick's chances can be predicted accordingly.

It has been shown for a variety of intuitions in mathematics and science that, with learning and education, what has seemed meaningless for a beginner can become a recognizable meaningful pattern for the expert. Dennett (1991) elaborates the idea that differences in knowledge yield striking differences in people's capacity to pick up patterns. Studies of chess experts (de Groot, 1965) have shown that expertise enables an immediate recognition of an apparently complex pattern, and this recognition, in turn, enables a rapid and accurate response. This capacity is usually called "intuition", just as the physicist's rapid response to questions in physics is considered physical intuition (Larkin, McDermott, Simon, & Simon, 1980). Kahneman and Tversky (1982), who have studied statistical intuitions, contend that these intuitions vary with intelligence, experience, and education. As in other domains of knowledge, what is intuitive for the expert is often non-intuitive for the novice.

New secondary intuitions can be acquired in the process of learning. Becoming an expert involves developing intuitions of the initially non-intuitive (see Fischbein, 1987; Shaughnessy, 1991). The quality of teaching can be judged by the extent to which students manage to assimilate new principles. When a principle is well understood and is in harmony with a set of our beliefs and anticipations, it can be internalized to the point of becoming almost intuitive. Such could be the case with the criterion of comparing $P(h \mid D)$ with the weighted average of the other likelihoods and judging, by the results of that comparison, the direction of change in Dick's probability. I am so convinced of the soundness of that prescription by now, that I believe I have known it all along.

Some didactic comments

Which hat did you draw that observation out of?

An idea that is interwoven throughout the discussion of the three-prisoner problem is that we need to consider the *chance set-up*. That phrase, coined by Hacking (1965, p. 13), refers to what standard probability textbooks regard as the *statistical experiment*. The chance set-up is the setting and the random process that has generated the data of our probability problem. As we saw, it makes a difference whether the warden names Harry because the outcome of a toss-up has been H, or because he is biased toward naming Harry (meaning that some other "random device", whose probability of turning H is much higher, has yielded that outcome).

The three-prisoner problem is prototypic of a class of probability "paradoxes" that require a careful analysis of the possible origins of the information received

(Zabell, 1988a). A failure to provide a precise and unequivocal definition of the experimental procedure involved is a common source of confusion and error in many problems dealing with chance (Gardner, 1961). What matters for reaching a correct solution to many probability problems is often not only the given information, but also the manner by which it has been obtained (Bar-Hillel & Falk, 1982; Faber, 1976; Falk, 1983; Falk & Konold, in press; Shimojo & Ichikawa, 1989).

Imagine, for example, that the same evidence as in the original problem is now given by a warden who abides by another agreement. He is to flip a coin. If the outcome is H, he is to report *Harry's fate*, whatever it is. If the outcome is T, he reports Tom's fate. In this case, the very same statement "Harry will be executed" (denoted h) will have a different impact on Dick's prospects. Under this chance set-up $P(h \mid D) = 1/2$ as well as $P(h \mid T) = 1/2$ (as before, $P(h \mid H) = 0$). The weighted-average criterion predicts an increase in the target probability (since $P(h \mid D) > 1/4$), and Dick's survival probability indeed rises to 1/2 (as can be verified by inserting these values into Bayes' formula).

Describing the mechanism that has generated our observations may sometimes require more than maintaining that certain proceedings have been carried out "randomly". The underlying random model has to be fully specified. Its role may be crucial. Suppose you are to distribute *at random* r particles among n cells. It turns out that several methods may comply with that instruction. Offhand, it would seem that all n^r arrangements of r distinguishable objects in n cells, without any limitation, should be equally likely. But, two other interpretations (at least) of the randomness requirement come to mind: maybe the "individuality" of the particles is immaterial and one should attain equiprobability of all $\binom{n+r-1}{r}$ arrangements of r indistinguishable objects in n cells (without restrictions). A third possibility is to add a limitation of one particle, at most, per cell ($r \leq n$ in this case), and to interpret the task as that of selecting r cells from a total of n, so that all $\binom{n}{r}$ arrangements would be as probable.

In statistical mechanics it is common to subdivide the phase space into a large number, n, of small regions or cells and to describe the state of the entire system in terms of a random distribution of the r particles in n cells (Feller, 1957, pp. 38–40; Troccolo, 1977). The above three random models, which give rise to different combinatorial formulae, are known, in turn, as *Maxwell–Boltzmann's*, *Bose–Einstein's*, and *Fermi–Dirac's*. The important point, for the sake of our discussion, is that various chance procedures, all pertaining to the same physical set-up, may be equally legitimate. None of them is a more justified probability model on a priori grounds.[9]

[9]In reality, physical particles were never found to behave in accordance with Maxwell–Boltzmann statistics. It was shown that Bose–Einstein statistics apply to photons, nuclei, and atoms containing an even number of elementary particles, while the Fermi–Dirac model fits the distribution of electrons, neutrons, and protons (Feller, 1957).

In problems of geometric probability, a given short-hand instruction to perform "randomly" must be viewed with caution because it can be fraught with ambiguity (Gardner, 1961, pp. 221–226). A typical example is that of *Bertrand's Paradox*, published in 1889 (Lacey, 1962; Székely, 1986, pp. 43–48). It asks about the probability that a chord drawn at random inside a circle will be larger than the side of an equilateral triangle inscribed in the circle.

Several legitimate procedures of obtaining a "random chord" can easily be devised, each of which entails a different answer. One method suggests choosing a point uniformly from the entire area within the circle. This point is the midpoint of a uniquely determined chord. The method results in an answer of 1/4 (see Gardner, 1961; Székely, 1986). But one may also choose a random point uniformly on a radius of the circle and take the chord that is perpendicular to the radius at that point. This method entails an answer of 1/2. Other methods, no less "natural", may lead to 1/3 and perhaps to other probabilities (Lacey, 1962).

These different results are paradoxical only if one believes that a "random chord" is a well-defined concept. Since each of several procedures is a legitimate method of obtaining a "random chord", the problem, as originally stated, is ambiguous. It has no answer until the meaning of drawing a chord at random is made precise by a description of the procedure to be followed.

Similar problems and paradoxes have been discussed by Shafer (1985). He develops the idea of insisting on the statement of a *protocol*, which amounts, in fact, to specification of all the contingencies of the statistical experiment. Protocols are important because we can properly interpret new information only when we know the rules governing its acquisition. A protocol that tells, at each step, what the warden might do next (and with what probability), is needed in the three prisoners' case to make conditioning on the new information legitimate. In a similar vein, Nisbett and Ross (1980) advise teachers to offer useful slogans for didactic purposes. One of their slogans, *Which hat did you draw that sample out of?* (p. 283), may be instrumental in alerting students to reflect on the chance mechanism underlying the problem at hand. Freedman, Pisani, and Purves (1978) anticipate that advice: they keep reminding the student, throughout all the probabilistic discussions in their textbook, to formulate the problem situation in terms of a *box model*.

Recently, several authors pointed out that neglect of proper explication of the box model, or the chance set-up, might be partly responsible for the base-rate neglect found in the research on heuristics and biases, as, for example, in Kahneman and Tversky's (1973) *engineer–lawyer problem* and in Tversky and Kahneman's (1980) *cab problem*. There was a brief mention that the experimental descriptions were "chosen at random" from a given population in the original engineer–lawyer problem, and no mention of any sampling procedure concerning the target cab in the cab problem. However, when Ginossar and Trope (1987) presented the same problems framed as games of chance, the same information

yielded robust base-rate effects. By vividly describing the sampling procedure that ended in observation of the target stimulus, the experimenters made the base-rate information more applicable. Gigerenzer, Hell, and Blank (1988) went one step further: they made the crucial condition of random sampling (in the engineer–lawyer problem) visually observable. The subject actually drew a description randomly from an urn. Their results showed that the base-rate neglect phenomenon was much reduced. The lesson suggested by these studies is that if experimenters (and subjects) who study statistical intuitions highlight the explication of the (random) process that generated the data, some of the well-documented fallacies may disappear.

Drawing analogies

The detailed analysis of one particular problem, with all its potential pitfalls, would be worthwhile if we can draw lessons that may be extended beyond the problem's specific circumstances. The three-prisoner problem, indeed, seems to encapsulate many of the cognitive hazards of probability problems, and the heuristics employed in dealing with it are characteristic of major intuitive strategies in problem solving under uncertainty. Furthermore, the advances gained in resolving this one problem can readily be generalized to a wide range of problems. As we saw, the caveat not to take the seemingly "given" information at face value but to inquire about the mechanism by which it has been acquired, may provide the key for understanding the differences among some physical phenomena and for solving geometric probability problems. Let us consider another example, from the domain of biology (see Falk, 1989 for more details).

A woman is expecting twins. A priori, the three possible combinations (two boys, two girls, boy and girl) are known to be equiprobable.[10] A chromosomal test is performed on cells sampled by chance from one (random) amnion, and the results show it is a boy. What is the probability that the woman is expecting two boys? Although the analogy with the three-prisoner story is not immediately obvious, one can by now sense that besides the fact that two girls are ruled out by that test, the remaining two possibilities are no longer equally probable. The chromosomal diagnosis of a boy (like the warden's statement that Harry is to be

[10]When considering any two independent births, including those of (nontwin) siblings, the probabilities of two boys or of two girls are each 1/4, and the probability of a boy and a girl is 1/2. Twins, however, can be either monozygotic (identical), in which case they are necessarily of the *same sex*, or dizygotic (fraternal), in which case their sexes segregate independently. The net result of the pooling of these two kinds of twins is that the distribution of pairs of twins differs from that obtained in the case of independence. Empirically, in the U.S.A., the three possible pairs of twins (two boys, two girls, boy and girl) were found to be about equally frequent. This can be shown to imply that about one out of three sets of twins born are identical (Stern, 1960, pp. 533–534).

executed) is impossible when assuming two girls (when Harry is to be pardoned); it is certain if conditioned on two boys (if Tom is to be pardoned); and it is a toss-up if conditioned on boy girl (if Dick is to be pardoned). No wonder that the a posteriori probability of two boys turns out to be 2/3 (just like $P(\mathrm{T}\,|\,\mathrm{h})$), and that of boy girl is 1/3 (like $P(\mathrm{D}\,|\,\mathrm{h})$).

Many other probability problems, some of them well known, like Monty's dilemma, are made up of similar ingredients, and their resolution involves the same line of thought. To name just a few, such is the case with the "second ace" problem, where one is asked about the conditional probability that two cards, randomly dealt to a player, are both aces, given that one is an ace. The analysis of that problem hinges on the specific features of different scenarios, that end by our learning of the existence of one ace in the player's hand (Faber, 1976; Freund, 1965; Shafer, 1985). Likewise, the "three-card problem" (Bar-Hillel & Falk, 1982) and various puzzles about two-children families (Gardner, 1961; Glickman, 1982) all manipulate the same source of ambiguity.

The degree of success with which one can apply a method of solution of one problem to another problem depends, of course, on the nature and extent of the analogy between them. Analogies are often based on a host of assumptions. These should be examined carefully so as to establish whether a particular analogy is appropriate or whether it may mislead us (Bransford & Stein, 1984, pp. 73–74 and p. 88). Analogy pervades all our thinking. It is used on very different levels of precision (Polya, 1957, pp. 37–46). Vague, incomplete, or unclarified analogies may fail the problem solver who relies on them. But analogy may reach the level of mathematical precision. At best we may have a one-to-one correspondence between the objects of two systems, which preserves certain relations. That is, if a certain relation holds among the objects of one system, the same (or an analogous) relation holds among the corresponding objects of the other system. Such an analogy is called an *isomorphism*. It is an information-preserving transformation.

We may consider ourselves lucky when, trying to solve a problem, we succeed in discovering a highly analogous problem which is simpler in some sense. Monty's dilemma is, in fact, isomorphic to the three-prisoner problem, but it has an extra twist: one can decide to switch, namely, to change the choice of a door (unlike Dick, who cannot swap fates with Tom). The questions addressed to the solver are somewhat different in these two cases. In the three-prisoner situation one is asked about any potential change in Dick's survival probability, whereas in Monty's game the question is whether the probability of the unopened door had risen above 1/2 so as to warrant a switch. Many resourceful efforts have been directed at clarifying Monty's problem, because of the publicity of the discussion around it. Let's look into two of these attempts and see whether, by analogy, they may enlighten us with respect to the three prisoners.

When trying to convince the skeptics that the contestant should switch to door

no. 2, after Monty has opened door no. 3, C. Konold (personal communication, February 19, 1991) uses the following argument: suppose that from the beginning your choice was between (a) opening door no. 1 and (b) opening doors no. 2 *and* no. 3. You would have chosen (b) for sure because this gives you a 2/3 chance of winning the prize. Therefore, if you adopt the decision to switch, no matter which door Monty opens, you are, in fact, preferring the other two doors to the one of your original choice. The only way for you to win by sticking to your original choice is if you have chosen the correct door to start with, and, as we know, the probability of that event is 1/3.

Note that the above reasoning has totally ignored Monty's procedure for deciding which door to open when he has a choice. A careful analysis reveals that, even if Monty is biased in favor of opening one particular door, but you don't know the direction of that bias, it would be advisable for you to switch. Suppose Monty will always open door no. 3 [when he can. If we, who know about that bias, observe him opening door no. 3], we would not see any reason to switch, because doors no. 1 and no. 2 are now equally likely to hide the prize. If, however, Monty opens door no. 2, we would know that the prize is certainly behind door no. 3 and we would switch. Overall, averaged across these possibilities, if you don't know the direction of Monty's bias, you should switch.[11]

The implications for our unfortunate prisoner, Dick, are quite gloomy. Whether the warden flips a coin to decide between naming Tom or Harry (whenever a choice exists), or whether he is biased in favor of naming one of the two, but Dick doesn't know which one, the information the warden gives doesn't change Dick's chances of pardon. But if offered the opportunity to switch his fate for that of the unnamed prisoner, he ought to jump at the chance.

A second strategy for clarifying the situation – one which often proves helpful for problem solving – is considering an extreme case. It is applied to Monty's dilemma by Shaughnessy and Dick (in press). (Imagining increasing the number of doors is much easier than imagining a million prisoners condemned to death . . . .) Gardner (1961) tries that strategy on a card problem devised to be isomorphic to the three-prisoner problem. Let's look at Marilyn's exaggerated version:

> Suppose there are a *million* doors, and you pick door No. 1. Then the host, who knows what's behind the doors and will always avoid the one with the prize, opens them all except door No. 777,777. You'd switch to that door pretty fast, wouldn't you? (*Parade* magazine. February 17, 1991, p. 12).

[11]Monty's bias toward opening door no. 3 might be partial. Let us assume that Monty will open door no. 3 with probability p (where $0 \leq p \leq 1$) when the prize is behind door no. 1. By Bayes' theorem one obtains that in such a case the posterior probability P that door no. 2 hides the prize, after observing a goat behind door no. 3, is $P = 1/(1 + p)$, which means that the probability of winning upon switching satisfies $1/2 \leq P \leq 1$. It would therefore be wise for the contestant to switch (Gillman, 1991, 1992).

There can hardly be a more smashing argument. Only that its truth depends on the understanding that if the prize is behind door no. 1, the host decides by a fair draw which one of the remaining 999 999 doors to leave closed. If, however, you know that for some reason or other the host is determined to leave door no. 777 777 closed, whenever possible, observing that situation will render that door as likely to hide the prize as door no. 1. You would then have no reason to hurry to switch.

The same is true for Dick. It all depends on what Dick knows about the warden's behavior. Would the warden randomly name one of the two when facing a choice, or would he prefer naming Harry (Tom)? This closes the circle by bringing us back to the starting point, that is, to the proper definition of the statistical experiment.

References

Bar-Hillel, M., & Falk, R. (1982). Some teasers concerning conditional probabilities. *Cognition*, *11*, 109–122.

Beckenbach, E.F. (1970). Combinatorics for school mathematics curricula. In L. Råde (Ed.), *The teaching of probability and statistics* (pp. 17–51). Stockholm: Almqvist & Wiksell.

Bransford, J.D., & Stein, B.S. (1984). *The IDEAL problem solver*. New York: Freeman.

de Groot, A.D. (1965). *Thought and choice in chess*. The Hague: Mouton.

Delbrück, M. (1986). *Mind from matter? An essay on evolutionary epistemology*. Palo Alto: Blackwell.

Dennett, D.C. (1991). Real patterns. *Journal of Philosophy*, *88*, 27–51.

Diaconis, P., & Zabell, S. (1986). Some alternatives to Bayes's rule. In B. Grofman & G. Owen (Eds.), *Proceedings of the second University of California, Irvine, conference on political economy* (pp. 25–38). Greenwich, CT: JAI press.

Faber, R.J. (1976). Re-encountering a counter-intuitive probability. *Philosophy of Science*, *43*, 283–285.

Falk, R. (1978). *Problems in probability and statistics*, Part 2 (in Hebrew), Jerusalem: Akademon.

Falk, R. (1983). Experimental models for resolving probabilistic ambiguities. *Proceedings of the Seventh International Conference for the Psychology of Mathematics Education* (pp. 319–325). Israel.

Falk, R. (1989). Inference under uncertainty via conditional probabilities. In R. Morris (Ed.), *Studies in mathematics education. Vol. 7: The teaching of statistics* (pp. 175–184). Paris: Unesco.

Falk, R., & Konold, C. (in press). The psychology of learning probability. In F.S. Gordon & S.P. Gordon (Eds.), *Statistics for the twenty first century*. USA: Mathematical Association of America.

Feller, W. (1957). *An introduction to probability theory and its applications*, 2nd edn. (Vol. 1). New York: Wiley.

Fischbein, E. (1987). *Intuition in science and mathematics: An educational approach*. Dordrecht: Reidel.

Freedman, D., Pisani, R., & Purves, R. (1978). *Statistics*. New York: Norton.

Freudenthal, H. (1970). The aims of teaching probability. In L. Råde (Ed.), *The teaching of probability and statistics* (pp. 151–167). Stockholm: Almqvist & Wiksell.

Freund, J.E. (1965). Puzzle or paradox? *American Statistician*, *19* (4), 29 and 44.

Gardner, M. (1961). *The second Scientific American book of mathematical puzzles and diversions*. New York: Simon & Schuster.

Gardner, M. (1992). Probability paradoxes. *Skeptical Inquirer*, *16*, 129–132.

Gigerenzer, G., Hell, W., & Blank, H. (1988). Presentation and content: the use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 513–525.

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

Gillman, L. (1991, June–July). The car-and-goats fiasco. *Focus: The Newsletter of the Mathematical Association of America*, *11* (3), 8.

Gillman, L. (1992). The car and the goats. *American Mathematical Monthly*, *99*, 3–7.

Ginossar, Z., & Trope, Y. (1987). Problem solving in judgment under uncertainty. *Journal of Personality and Social Psychology*, *52*, 464–474.

Glickman, L.V. (1982). Families, children, and probabilities. *Teaching Statistics*, *4*, 66–69.

Glickman, L.V. (1986). *Problems, paradoxes and misconceptions: Justification for teaching the history of probability*. Unpublished manuscript. City of London Polytechnic, Faculty of Computing, Mathematics and Allied Studies.

Glickman, L. (1989). Why teach the history of probability? *Teaching Statistics*, *11*, 6–7.

Hacking, I. (1965). *Logic of statistical inference*. Cambridge: Cambridge University Press.

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

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

Kahneman D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*, 237–251.

Kahneman, D., & Tversky, A. (1982). On the Study of statistical intuitions. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: heuristics and biases* (pp. 493–508). Cambridge: Cambridge University Press.

Konold, C., Lohmeier, J., Pollatsek, A., Well, A., Falk, R., & Lipson, A. (1991). Novices' views on randomness. *Proceedings of the Thirteenth Annual Meeting of the International Group for the Psychology of Mathematics Education. North American Chapter*, *1*, 167–173. Blacksburg, VA.

Lacey, O.L. (1962). The human organism as a random mechanism. *Journal of General Psychology*, *66*, 321–325.

Larkin, J., McDermott, J., Simon, D.P., & Simon, H.A. (1980). Expert and novice performance in solving physics problems. *Science*, *208*, 1335–1342.

Morgan, J.P., Chaganty, N.R., Dahiya, R.C., & Doviak, M.J. (1991). Let's make a deal: the player's dilemma. *American Statistician*, *45*, 284–287.

Mosteller, F. (1965). *Fifty challenging problems in probability with solutions*. Reading, MA: Addison-Wesley.

Nisbett, R., & Ross, L. (1980). *Human inference: strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.

Phillips, L.D. (1973). *Bayesian statistics for social scientists*. London: Nelson.

Polya, G. (1957), *How to solve it*, 2nd edn. Princeton, NJ: Princeton University Press.

Saunders, S.C. (1990, April). The shell game, prisoner paradoxes of conditional probability. *Mathematics Notes from Washington State University* (Vol. *33*, No. 2, Whole Number 129).

Selvin, S. (1975a). A problem in probability. In Letters to the Editor. *American Statistician*, *29*, 67.

Selvin, S. (1975b). On the Monty Hall problem. *American Statistician*, *29*, 134.

Shafer, G. (1985). Conditional probability. *International Statistical Review*, *53*, 261–277.

Shaughnessy, J.M. (1991). Research in probability and statistics: reflections and directions. In D. Grouws (Ed.), *Handbook on research in mathematics education*. New York: Macmillan.

Shaughnessy, J.M., & Dick, T. (1991). Monty's dilemma: should you stick or switch?. *Mathematics Teacher*, *84*, 252–256.

Shimojo, S., & Ichikawa, S. (1989). Intuitive reasoning about probability: theoretical and experimental analyses of the "problem of three prisoners". *Cognition*, *32*, 1–24.

Stern, C. (1960). *Principles of human genetics*, 2nd edn. San Francisco: Freeman.

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

Troccolo, J.A. (1977). Randomness in physics and mathematics. *Mathematics Teacher*, *70*, 772–774.

Tversky, A., & Kahneman, D. (1980). Causal schemas in judgments under uncertainty. In M. Fishbein (Ed.), *Progress in social psychology* (Vol. 1, pp. 49–72). Hillsdale, NJ: Erlbaum.

Weintraub, R. (1988). A paradox of confirmation. *Erkenntnis*, *29*, 169–180.

Zabell, S.L. (1988a). The probabilistic analysis of testimony. *Journal of Statistical Planning and Inference*, *20*, 327–354.

Zabell, S.L., (1988b). Symmetry and its discontents. In B. Skyrms & W.L. Harper (Eds.), *Causation, chance, and credence* (Vol. 1, pp. 155–190). Dordrecht: Kluwer.