

Are People Programmed to Commit Fallacies? Further Thoughts about the Interpretation of Experimental Data on Probability Judgment

L. JONATHAN COHEN

I

THE NORM EXTRACTION METHOD

During the last decade or so a characteristic method of interpreting experimental data has been followed by most researchers in their investigations into how statistically untutored subjects actually judge probabilities. For the most part the investigator tacitly assumes that the problem-task set to his or her subjects is correctly soluble only in terms of some academically well-regarded conception of probabilities that he or she has in mind. The investigator therefore evaluates the subjects' performance for correctness or incorrectness by a technique of assessment or estimation that is appropriate to this mode of conception. If some subjects' responses seem, when thus evaluated, to be incorrect, their error may be put down by the investigator—in terms of the computational metaphor—either to a standing fault in the programming of ordinary people (in that they are not programmed to apply appropriate valid principles to the task in question or are programmed to apply inappropriate or invalid ones) or to a temporary malfunction in the running of a valid programme.

We may call this method of interpreting experimental data about probability judgments "the Preconceived Norm Method" and contrast it with what is conveniently called "the Norm Extraction Method". The Norm Extraction Method assumes instead that, unless their judgment is clouded at the time by wishful thinking, forgetfulness, inattentiveness, low intelligence, immaturity, senility, or some other competence-inhibiting factor, all subjects reason correctly about probability: none are programmed to commit fallacies or indulge in illusions. So by setting subjects a particular kind of problem-task the investigator may at best expect to discover from their responses what conception of probability they implicitly adopt for solving that kind of problem and how they assess or estimate the size of the probability.

In short, the Preconceived Norm Method assumes a standard conception of probability, imputes its acceptance to the subjects, and hypothesises either faulty programming or temporary causes of malfunction in order to account for estimates that are erroneous in terms of that conception: the Norm Extraction Method hypothesises about the subjects' conception of probability and their mode of assessing it, on the assumption that unless affected by temporary or adventitious causes of error their judgments are correct.

The issue between these two methods is not of a kind that lends itself to being determined by a single crucial experiment since it concerns a whole policy or paradigm of research. The principles implicitly involved are not just straightforward generalisations that are open to being refuted by the discovery of counter-instances. Instead they are multiply quantified claims which assert that *all* recurrent patterns of error in lay judgments of probability have *some* explanation of such-or-such a kind (in the one case, of either the faulty-programme kind or the cause-of-malfunction kind; in the other case, only of the cause-of-malfunction kind). *All-and-some* principles of this nature often play an important methodological role in the development of scientific research, but because of their logical structure they are inherently neither verifiable nor falsifiable (Watkins, 1958). The success of such a principle may be gauged only by the extent to which it helps to make coherent sense of the already available data throughout the relevant field of investigation and also continues to promote adequate explanations there. So it is essential to consider the issue between the Preconceived Norm Method and the Norm Extraction Method in the context of a wide range of different experiments. Thus in Cohen (1979) and Cohen (1981a) several examples were given of how allegedly prevalent fallacies about sample-size and prior probabilities may be seen not to be fallacious at all when treated in accordance with what has here been called the Norm Extraction Method: the subjects' allegedly erroneous responses in these cases turn out to be capable of being understood in a way that admits their validity. And in the present paper a number of other well-known results will be discussed in order to show how extensively the Norm Extraction Method fits the facts. It is now rather doubtful whether any firm evidence will ever be found that human beings are programmed to judge probabilities faultily in certain circumstances (e.g. by the so-called heuristic of representativeness).

Certain *a priori* considerations also favour this conclusion.

One such consideration is that the Norm Extraction Method is more economical in the variety of explanatory factors that it envisages. It admits only normatively correct mechanisms on the one side, and adventitious causes of error on the other. The Preconceived Norm Method admits not only both these types of factor, but also certain invalid heuristics—cognitive routines that sometimes generate correct answers and sometimes do not. This is because, where subjects' responses do not conform to the assumed norm

but seem too regular and clearly patterned to be put down to adventitious causes, the Preconceived Norm Method tempts an investigator to suppose that these responses are programmed and therefore to postulate a special category of faulty programmes in order to account for them.

A second consideration stems from the fact that the mathematical calculus of chance has long since been recognised to have more than one possible interpretation as a theory of probability. As Hacking (1975, p. 13) points out, Poisson and Cournot saw, long before Carnap, that this calculus could be understood as providing principles for both an objective and an epistemic conception of profitability. Nor, in fact, are only two different interpretations admissible, as Nagel (1955) has shown. Rather we have on our hands a new problem: why are at least five strikingly different interpretations possible? For we can interpret the probability-function that is regulated by the classical calculus of chance either as a ratio of *a priori* chances, or as a relative frequency of empirically given outcomes, or as a causal propensity, or as a logical relation, or as a measure of actual or appropriate belief. Mackie (1973, p. 155ff) sought to explain this in terms of a series of historically natural shifts and extensions in the meaning of the term 'probability'. But this explanation is not only unsupported by the chronological facts. It is also inadequate to the demands of the problem (Cohen, 1977, pp. 11–12). Instead we have to think of probability as a generalisation of the concept of provability and as necessarily then taking a variety of semantically specific forms in accordance with familiar distinctions between different kinds of proof-systems (Cohen, 1977, p. 13ff.). Moreover, if for that reason any coherent gradation of provability is a probability, we must also allow the possibility of certain non-Pascalian kinds of probability-judgments which do not even conform to all the constraints imposed by the classical (Pascalian) calculus of chance. In particular we can have non-additive probabilities—probabilities that do not conform to a complementational principle for negation—as Bernoulli long ago recognised (Shafer, 1978) and as the Baconian tradition also held (Cohen, 1980). Just as we can usefully measure quantities of apples in at least three different ways (by number, by weight or by volume), so too, if we so desire, we can measure probabilities in several different ways according to the purpose in hand. *A* may well be more probable than *B* by one measure, and less probable by another.

Since therefore more than one conception of probability is possible, an experimenter has somehow to establish which is operative in the reasonings of his subjects about the problem-task presented to them. But how can he do this independently of the actual data that emerge from the experiment? If he recognises the risk of ambiguity and instructs the subjects beforehand in which conception of probability they are to use and how they are to use it, he is no longer dealing with naive and untutored subjects. The subjects would have been tutored in a certain conception of probability, and any

errors that they then commit may legitimately be treated as a measure of the inadequacy of the experimenter's instructions rather than as a measure of their own native irrationality. So, if the experimenter sets his subjects a real-world task and does not instruct them beforehand about which conception of probability they are to use, he has to extract from their behaviour what conception they actually use. And this has to be done within a total scheme of interpretation that makes due allowance for the possibility that subjects may make genuine errors in estimating the sizes of particular probabilities that satisfy their conception. Certainly it is unsafe for the investigator to rely on having formulated the wording of the task so unambiguously that it is bound to convey to the mind of each untutored subject precisely the same problem as the investigator himself has in mind (cf. Ebbesen and Konečni, 1980, p. 38, and J. S. Carroll, 1980, p. 70).

A third consideration is also worth mentioning, though opinions may differ about how far it takes us. Where deductibility hinges on the meanings of logical particles or quantifiers, ordinary people cannot be supposed to be systematically incorrect in their intuitions about what is deducible from what. The criteria of correct deduction are implicit in their own speech practice—although what has to be taken into account in each case is the total content of the message communicated and not just its linguistic form (Cohen, 1981a, p. 326f.). Of course, people sometimes contradict themselves or make invalid deductions. But the norms that they infringe when they do so are discoverable from their own intuitions. So to understand them correctly one must for the most part be charitable, and assume that they have not spoken illogically. How far can an analogous point be made about probability? The trouble here is that one needs to distinguish between this or that conception of probability, on the one hand, and appropriate strategies for assessing or estimating it, on the other.¹ For example, when probability is conceived as a relative frequency, a variety of sampling techniques are available for its estimation, and when probability is conceived as a measure of the strength of justified belief, there is a variety of ways in which coherent schemes of betting-odds may be devised in order to grade this strength in a particular context. And in each of these two cases the conception of probability that is operative can, at least in principle, be identified independently of the method of estimation employed.

But it does not have to be so identified. That is to say, there are two ways of individuating the operative conception of probability. One can think of it rather narrowly, e.g. as a relative frequency to be estimated by such-and-such a sampling strategy. Or one can think of it, more broadly, as a relative frequency however estimated. In terms of the former, narrower mode of individuation, a subject's conception of probability has to be inferred (subject to the possibility of adventitious error) from the nature of the subject's own responses. Hence there can be no room to impute erroneous

programming. But even if this strongly apriorist argument is unacceptable, and the broader mode of individuating conceptions of probability is therefore preferred, so that the fundamental question at issue becomes an empirical one, the evidence nevertheless tends in the same direction. It is logically possible, for example, that the ordinary layman supposes small samples to be more indicative than large ones in relation to long-run relative frequencies. But it turns out empirically false that he does so (see sec. VII below).

Charitableness, however, has only a rather limited role to play in the interpretation of experimental data about probability judgment. It would be a highly misleading caricature of the Norm Extraction Method to represent it as a policy of always, or nearly always, interpreting subjects' responses in a sense in which they are correct. Such a policy would indeed be methodologically indefensible: we are not entitled to infer *just* from subjects' methods of assessing or estimating probabilities to their use of a conception of probability for which these are reasonable answers. It is worth emphasising, therefore, that two constraints exclude the Norm Extraction Method from permitting this procedure. First, we have to suppose that the conception of probability in terms of which a subject comes to construe his task is cued for him in each case by the wording and content of his instructions (against the background of his own experience and education and any other individual differences that are relevant to his cognitive performance). Hence the Norm Extraction Method requires investigators to hypothesise regularities in such cueing which block arbitrarily charitable attributions of validity. Relevantly similar subjects must be expected to operate with the same conception of probability in relevantly similar contexts. Secondly, people do make mistakes quite often, and the causes of error must also be supposed to function regularly. So if an error is imputed to a subject in one case and attributed to a particular cause of malfunction there in order to provide a coherent account of the situation, then that factor must be expected to cause similar malfunctions in all relevantly similar cases.

II

THE CONCEPT OF SUBJECTIVE PROBABILITY

Research-workers in psychology, as in other biological sciences, most naturally think of probabilities as long-run relative frequencies. Such probability-functions take sets as arguments and their values are ratios between the cardinalities of those sets. In this way psychologists can see themselves as dealing with inherently measurable quantities and can thus seek to maintain the characteristically quantitative objectivity of modern science. It would hardly be surprising, therefore, if some of them were unselfcritically inclined to suppose that laymen also think most naturally of probabilities in

the same terms. But ordinary people, it appears (see sections III-VII below), do not naturally think of probability in such terms unless appropriately cued to do so. They either think of probability within a non-Pascalian framework, or they prefer a Pascalian probability-function that takes propositions as arguments, in accordance with the so-called "subjectivist" or "personalist" account, or one that takes properties as arguments, in accordance with the "propensity" account. And here it is necessary to point out that the term "subjective probability" is ambiguous, and ambiguous in more than one way.

The first ambiguity arises from a difference between psychological and philosophical usage. It is widely assumed by psychologists that "the subjective probability of X ... is veridical to the extent that its value agrees with the corresponding objective probability" (Peterson et al., 1965, p. 526). Now that assumption is certainly correct, though trivial, where "the subjective probability of X " means "a subject's estimate of the objective probability of X ". But it is just as certainly incorrect where "the subjective probability of X " means "the strength of belief in X " and "the objective probability of X " means "the relative frequency of X ". When people measure the probability of an individual outcome by evaluating the strength of belief that they have in that outcome in accordance with acceptable odds on its occurrence, as proposed by Ramsay (1931), De Finetti (1931), Savage (1954), and Kyburg and Smokler (1964), they can take into account not only what they know about relevant relative frequencies but also what they know about anything else, including the extent of their knowledge about relevant relative frequencies. The probability that Jones will survive to the age of 70, on the evidence that he is a university teacher, depends then not just on the relative frequency of such survivals among university teachers but also on the extent to which other relevant information about Jones (his medical history, leisure activities, etc.) has not been included in the evidence. The odds on Jones's dying before the age of 70 might be quite long if the only relevant fact that needed to be examined were his mode of earning a living and he was known to be a university teacher. But if only that fact were known, whereas there were several other relevant questions about Jones that remained uninvestigated, a bookmaker would shorten his odds in order to reduce his risk. The possibility that, in this sense, subjective probabilities—or "personal probabilities", as it will be less confusing to call them—may differ legitimately in value from the corresponding objective relative frequencies turns out to be vitally important for the interpretation of experimental data on probability judgment.² Indeed, in view of the fact that many Bayesian theorists (Savage, 1964) have combined their preference for Bayesian modes of calculation with a preference for the personalist conception of what they are calculating (or have even thought that the latter preference provided a rationale for the former), it is quite surprising to find that many psychol-

ogists are inclined to combine respect for Bayesian norms with a tendency to ignore the legitimacy of the personalist conception. Only thus has it seemed at all plausible for them to decry "the fallacy of conservatism in information integration", "the gambler's fallacy", "the fallacy of evaluating probabilities by sample-to-population ratio", and so on. Of course, it is possible that these psychologists all reject the possibility of measuring probabilities in the way suggested by the Ramsey-De Finetti-Savage school. But, though the measure in terms of coherent betting-odds may be inapplicable in certain kinds of cases, such as issues about the past where no more evidence can be obtained (Cohen, 1977, pp. 26-27, and 1981b), its mathematical possibility is scarcely in doubt. Or at any rate the rejection of that possibility needs to be supported by new argument, and no such support is to be found in the relevant psychological literature.

But there is also another important ambiguity in the notion of subjective or personal probability. When a person judges the probability of X by the strength of his belief in X, does his statement of X's probability mean just that he himself, on the evidence, believes in X with that strength of conviction? Or does it mean that anyone, on the evidence, ought to believe in X with that strength of conviction? The former (autobiographical) version of the personalist account has a feature that distinguishes it from all other analyses of probability judgment: two people who ascribe different probabilities to the same outcome are not contradicting one another, since each is merely describing his own mental state. But the other (deontic) version of the personalist account lacks this idiosyncratic feature and is correspondingly more widely applicable. Indeed, if it can be construed as measuring by appropriate ratios wherever betting odds are inappropriate, it could be thought of as differing from the analysis of probability as a causal propensity only when looked at from a rather metaphysical point of view.³ Wherever both analyses operate, both should assign the same probability-values to the same outcomes, since the propensity analysis allows us to take into account the extent to which the influence of Jones's profession on his survival might be affected by other unstated factors just as the personalist analysis allows us to take into account the extent to which other information about Jones might be relevant. But the propensity analysis will be more congenial to those who prefer to think that their judgments about probability are describing something that is objectively real, while the personalist account will be more congenial to those idealists, empiricists and others who prefer to confine the real to the observable. Fortunately the subjects in the experiments to be discussed need not be supposed to concern themselves with this metaphysical issue.

III

CONSERVATISM IN INFORMATION INTEGRATION

Let us now explore some of the consequences of applying the Norm Extraction Method instead of the Preconceived Norm Method. We shall see that the concept of personal probability, or its objectivist counterpart, has an important part to play.

Consider first the alleged fallacy of conservatism, which has been authoritatively claimed (Slovic, Fischhoff and Lichtenstein, 1977) to be "the primary finding of Bayesian information integration research". In certain easily reproduced circumstances, it is claimed, laymen do not make big enough revisions of their probability estimates when presented with additional information. For example, Phillips and Edwards (1966) had their subjects estimate whether a randomly chosen bag contained predominantly blue or predominantly red chips, by—in effect—drawing a chip at a time from the bag and replacing it. The subjects were told the ratio of bags containing predominantly blue chips to those containing predominantly red ones, and this was assumed to afford them a prior probability. They were also told the number of chips of the majority and minority colours, which were the same in each bag. A sequence of twenty chips was shown them, drawn one at a time from the chosen bag and each replaced immediately after being shown. And the subjects were asked to revise their previous intuitive estimates (of the probability that each type of bag had been chosen) after each new chip was shown. In the event the subjects revised these estimates less than the amount that the investigators calculated they should have done according to Bayes law. So the investigators concluded that conservatism in the integration of new information was a pervasive effect, increasing as the diagnostic value of a single chip increased. Similar results have been reported by several other investigators such as Peterson and Du Charmé (1967), Messick and Campos (1972) and Donmell and Du Charmé (1975), despite the fact that this allegedly excessive respect for prior probabilities seems difficult to reconcile with subjects' alleged tendency, reported by Kahneman and Tversky (1973), Lyon and Slovic (1976) and Bar Hillel (1980), to ignore prior probabilities altogether.

Is there a real fallacy here? Of course, a fallacy has to be admitted if we assume, in accordance with the Preconceived Norm Method, that the subjects are conceiving the probability of an outcome as the ratio of the number of favourable chances to the total number of possible chances. On that conception the relevant likelihoods—e.g. the probability of drawing a red chip (D), on the hypothesis that the bag contains 70 red chips and 30 blue ones (H)—are strictly determined by the terms of the problem. "Simple statistical theory" (Alker and Hermann, 1971, p. 134) suffices to settle these likelihoods. And the correct revisions are then immediately apparent from

an application of Bayes' theorem, since the prior probability of H has been stated and the prior probability of D can be calculated from the information given.

But the subjects are surely entitled to be a little cautious about just assuming that the selection of a chip of particular colour from a bag is a matter of pure chance. Who knows how the bag was filled or shaken, or what the effect this will have on drawings from it? "Simple statistical theory" may not be the right guide to follow here.⁴ The layman's black box should be expected to be as pedantic as a computer in not assuming input that it has not had. So let us pursue the Norm Extraction Method instead and seek an interpretation of the subjects' probability function that is more in keeping with their state of information and the content of their responses. Suppose that the subjects conceive probability as a measure of appropriate intensity of belief. They will not then treat the hypothesis about the bag's composition as being quite so relevant, whether positively or negatively, to the probability of picking a chip of a particular colour. $p(D, H)$ will therefore be rather closer to $p(D)$ than it would be on the assumption that pure chance is operating, so $p(D, H)/p(D)$ will be nearer to 1. Therefore by Bayes' theorem, which says that (where $p(D) > 0$)

$$p(H, D) = \frac{p(D, H) \times p(H)}{p(D)},$$

$p(H, D)$ will be closer to $p(H)$. In other words the fallacy of conservatism here is an artifact of the Preconceived Norm Method. If instead we apply the Norm Extraction Method, we can interpret the experimental data in a way that allows us to infer the subjects to have been consistently Bayesian in their calculations.

Admittedly Vlek and Van der Heijden (1967) claim to have found that for individual data, direct posterior probability estimates are not systematically related to values inferred from likelihood estimates. But this finding seems to be contradicted by results that Peterson et al. (1965), Beach (1966), Messick and Campos (1972) and Vlek (1973) describe.

Two other experimental findings help to support the proposed interpretation.

First, if so-called "conservatism" in the face of present evidence is really the result of conceiving probability as the right strength of belief rather than as the ratio of favourable chances, one would expect that any encouragement of subjects to conceive of probabilities in the latter terms would diminish or eliminate conservatism. And so it turns out (Messick and Campos, 1972): calibration in such cases destroys conservatism.

Secondly, if this so-called "conservatism" resulted from some inherent inadequacy in people's information-processing systems one might expect that, when individual differences in information-processing are measured on

independently attested scales, some of them would correlate with degrees of "conservatism". In fact no such correlation was found by Alker and Hermann (1971). And that is just what one would expect if "conservatism" is not a defect at all, but a rather deeply rooted virtue of the system.

Other experimental findings in this area seem too uncertain to be taken into account. For example, when data confirm a subject's hypothesis, "conservatism" is sometimes reported as being increased (Du Charme and Peterson, 1968) and sometimes as being decreased (Pitz, Downing and Reinhold, 1967). It is difficult to understand this difference unless it is some kind of experimental artifact. Similarly, subjects are said to get even more conservative as the perceived importance and complexity of their task increases (Alker and Hermann, 1971). But unfortunately in just these cases it becomes riskier to take for granted that the subjects' judgments of prior probabilities analogous to $p(D)$ or $p(H)$ above are psychologically independent of their judgments of likelihoods analogous to $p(D, H)$: "it is precisely the amount of independence that may be changing as problems become less like mathematical puzzles and more like 'real-life' decisions" (Alker and Hermann, 1971, p. 39). Finally, there is some evidence that "conservatism" may increase in later trials (Du Charme and Peterson, 1968) though this effect is not reported by Phillips and Edwards (1966). If the effect is genuine, it would just suggest an increasing reluctance on the part of subjects to assume that the data are being obtained by a random or chance process.

IV

THE GAMBLER'S FALLACY

The same manoeuvres—replacing a ratio-of-chances interpretation of the subjects' probability judgments by a personalist or propensity one—will serve to transform the so-called "gambler's fallacy" into a valid conclusion from a trio of premisses, each of which is inherently reasonable in the situation concerned. Investigators who ignore this (e.g. Tversky and Kahneman, 1971) do so at the risk of misinterpreting subjects' responses. Of course, where the probability of any single outcome in a game of chance is measured by the inverse of the number of severally exclusive and jointly exhaustive outcomes, it is a fallacy to suppose that the probability of one particular outcome on the present occasion is in any way affected by the actual outcome on any other occasion. If probability is measured in terms of *a priori* given alternatives, as is appropriate for undoubted games of chance, then *ex hypothesi* no physical event or empirical fact can possibly be relevant to the determination of the probability. For example, if we think of coin-tosses in these terms, then, however long a run of heads has occurred with a particular coin, it is a fallacy to suppose that the probability of the coin's falling tails on the next toss is anything less, or greater than, $\frac{1}{2}$. But the situation is

quite different for a person who does not take as certain that the game is one of pure chance and therefore measures the probability instead by appropriate intensity of belief or strength of natural propensity. For such a person it becomes reasonable to take into account any relevant empirical facts, such as the balance of the coin and its toss, or the long run of heads, and each of the following three premisses becomes inherently plausible. The derived conclusion then follows by a series of deductively valid steps. If the three premisses are true, the conclusion must also be true:

- Premiss (i): It is highly probable (i.e. much more probable than not) that every toss's outcome is a matter of chance. (I.e., the presumption is that the causal factors controlling the outcome of a toss cannot be determined sufficiently to found a prediction of the outcome.)
- Premiss (ii): if it is highly probable that every toss's outcome is a matter of chance, then, in the case of each toss's outcome, that outcome will probably not increase the present low probability that the coin, or the method of tossing it, is biased. (I.e., if you're strongly inclined to believe that the causal factors controlling the outcome of a toss cannot be determined sufficiently to found a prediction of the outcome, then you're inclined to believe of each future toss's outcome that it will not reduce your inclination to believe the causal factors controlling the outcome of a toss cannot be determined sufficiently to found a prediction of the outcome; or, more generally, if you're strongly inclined to believe that p , then you're inclined to believe of each as yet unknown piece of evidence that it will not be unfavourable to p .)
- Premiss (iii): if the next outcome, after the present long run of heads, is heads yet again, that outcome will increase the present low probability that the coin, or the method of tossing it, is biased. (I.e., despite the presumption in (i), only a born sucker—or a victim of the sorites paradox—never suspects a rigged game.)
- (iv): in the case of each toss's outcome that outcome will probably not increase the present low probability that the coin or the method of tossing of it is biased (by modus ponens from (i) and (ii)).
- (v): if the next outcome will not increase the present low probability that the coin, or the method of tossing it, is biased, then that outcome, after the present long run of heads, will not be heads yet again (by contraposition from (iii)).

(vi): if it is probable that the next outcome will not increase the present low probability that the coin, or the method of tossing it, is biased, then that outcome, after the present long run of heads, will probably not be heads yet again (from (v) in virtue of the rule "From X's being a consequence of Y infer $p(X) \geq p(Y)$ ").

Conclusion (vii): it is probable that the next outcome, after the present long run of heads, will not be heads yet again (by universal instantiation and modus ponens from (iv) and (vi)).

This argument may encounter the objection that any such subject who offered short odds on tails in accordance with the conclusion (vii) would be bound to lose money in the long run. But would he? The only apparent irrationality in such a bettor's conduct is that while offering short odds on tails he also seems committed to offering short odds on the truth of the proposition that evens are the correct odds on tails, since by (i) he assumes it probable that each toss's outcome is a matter of chance. And to live up to this latter commitment would in fact be a prudent and coherent way of laying off any bets made in accordance with (vii). A person who loses money in the long run on his bets in accordance with (vii) could win back his losses by betting, in accordance with (i), that he was accepting wrong odds in accordance with (vii). So the only way to challenge the argument seriously would be to prove that one or more of the premisses is fallacious. But that does not look like being an easy task. Certainly one should not think ill of a layman who asserts (vii), since his assertion can be rationalised as a deduction from inherently plausible premisses (even if he himself is unaware of this).

There may also be other ways of rationalising the gambler's fallacy (Cohen, 1981a, pp. 327–328), and no doubt many people get sufficiently muddled about probability to make genuine mistakes about it. In order to find out exactly what goes on, and when, we need more experimental data, with subjects' protocols. But there is at present no simpler potential explanation of the prevalence of the so-called gambler's fallacy than that for the most part it is not a fallacy at all and that all we can tell by a gambler's attitude of approval or disapproval towards the "fallacy" is whether he is measuring his probabilities by strength of belief (or natural propensity) or by *a priori* ratios, respectively. Not surprisingly, indeed, it is found that the negative recency effect exhibited in the gambler's fallacy tends to disappear after a large number of trials (Lindman and Edwards, 1961). That is to say, as calibration becomes possible premise (iii) above becomes less plausible.

Certainly the Preconceived Norm Method has not come up so far with any acceptable explanation. Those who assume that the probabilities concerned are *a priori* ratios, and that therefore the gambler's "fallacy" is a real error,

have to explain why people are so prone to commit this error. They are therefore led to attribute the error to the alleged prevalence of some fallacious heuristic, such as the so-called heuristic of representativeness. For example, Tversky and Kahneman (1971, pp. 105–106), followed by Nisbett and Ross (1980, p. 25), tells us that the gambler's fallacy is to be accounted for by such facts as that the occurrence of black on a roulette wheel after a long run of red is more representative of the generating process than the occurrence of another red. Yet the occurrence of black after only one occurrence of red, or of two blacks after two reds, would be even more "representative". So this theory leaves quite unexplained why it is that real gamblers are reputed to expect black after a long run of red but not after a short run, even though the latter would be more "representative" than the former. Nor are they reputed to expect a long run of black after a long run of red, though again this would be more representative. On the other hand, as we have seen above, if we follow the Norm Extraction Method these facts can be quite satisfactorily explained without our invoking the prevalence of any potentially fallacious procedure like the representativeness heuristic.⁵

V

JUDGING THE PROBABILITY OF A CONJUNCTION

Another profitable field for applying the Norm Extraction Method is in regard to experiments about judging the probability of a conjunction.

In a recent study Kahneman and Tversky (forthcoming, cf. Nisbett and Ross, 1980, p. 146) gave subjects personality profiles of various target persons. Subjects were then asked to assess the likelihood that the persons described in the profiles belonged to various groups. One group of subjects was asked to estimate the likelihood that persons with given profiles were members of non-compound groups like lawyers or Republicans. Another group of subjects was asked to estimate the probability that the profiled persons were members of compound groups like Republican lawyers. What Kahneman and Tversky found is that if a profiled person is judged rather unlikely to be, say, a lawyer, and rather likely to be a Republican, he will be judged moderately likely to be a Republican lawyer. So the investigators concluded that their subjects had committed the fallacy of supposing that the mathematical probability of a person's being a Republican lawyer is higher than that of his being a lawyer and they explained this by attributing to their subjects a heuristic of representativeness that was programmed in such a way as to commit this fallacy in such cases.

But, if we follow the Norm Extraction Method and assume instead that most subjects were not committing any fallacy here, it is easy enough to infer an appropriate interpretation of the data that pays due regard to how the subjects might be expected to view their task. All that is needed is to infer

that, in keeping with familiar findings of attribution theory, the subjects have taken the question to concern the causes or effects of the collections of features set out in the profiles and are out to assess the believability of a story about an individual that is cast in these terms, rather than to estimate the relative frequencies of different kinds of people. For it would be quite coherent to think of these features as rarely resulting just from a person's being a lawyer, very often from a person's being a Republican, and quite often from a person's being a Republican lawyer (because the effect of being a lawyer is to weaken, but not to destroy, the relevant causal efficacy of Republican Party membership). Or the features might quite coherently tend to encourage a person to become a Republican and to deter him from becoming a lawyer, unless he in fact becomes a Republican. Of these two possible attitudes towards the profiled features, the former seems inherently more plausible, since it treats the lawyer *versus* Republican question as a request for a diagnosis, as it were, of the symptoms present in the profile. But both attitudes can easily be reconstructed either within a personalist or propensity system of Pascalian probability or within a Baconian system of non-complementational probability that is appropriate for reasoning about causal connections (Cohen, 1979). So there is no need at all to suppose an intrinsically fallacious mechanism here.

Again, we are told (Nisbett and Ross, 1980, p. 147) that subjects estimate the probability of a compound sequence of events of a certain kind, on given evidence, to be greater than the least likely of the events in the sequence. But the interpretation appropriate for this datum is rather like that appropriate for alleged commissions of the fallacy-of-the-converse (Cohen, 1981a, pp. 326-327). One has to take into account not just the meanings of the sentences in which instructions to subjects are formulated but also the implications of uttering them. So when asked for the probability of a particular single event subjects may well infer that what is wanted is an estimate of the believability of that event's occurrence as an apparently isolated effect, which could well be lower than the believability of the occurrence of a particular causal sequence containing the event. If, on the other hand, subjects are asked specifically for the probability of the single event's occurrence *whether* in isolation *or* within the particular sequence, it would be very surprising indeed if they then went on to declare the particular sequence's occurrence to be even more probable.

The Norm Extraction Method also has an important bearing on that type of experiment with drawing marbles from urns in which significant numbers of subjects are alleged to overestimate the probabilities of conjunctive outcomes and underestimate the probabilities of disjunctive ones (e.g. Cohen and Hansel, 1957; Bar-Hillel, 1973). This interpretation of the data results from applying the Preconceived Norm Method, because it is assumed that all the subjects share the investigator's conception of the act of drawing a

marble out of a mixture of white and coloured marbles in an urn, with replacement, as a perfect game of chance. If that assumption were correct, then each act of drawing a marble (from an urn containing stated proportions of white and coloured marbles) would be thought independent of every other such act. And no doubt some of the subjects, who give what the investigators think to be the correct answers, do indeed conceive the act of drawing a marble in just this way. But the others, who give the allegedly incorrect answers, may according to the Norm Extraction Method be supposed to measure their probabilities in a way that allows their estimates to be correct. Specifically they may be supposed to measure their probabilities by the intensity of their beliefs and therefore, because of the measures they assign to compound events, to regard the draws as not wholly independent of one another. Can we be quite sure, they might be supposed to ask themselves (if they thought about it explicitly), that when a coloured marble is drawn first it may not be evidence of a tendency in the mixing process to bring coloured marbles within easier reach, or that when a coloured marble is replaced it may not tend to increase the number of coloured marbles that lie most readily to hand within the urn? Is not therefore the probability of drawing two coloured marbles slightly higher than the square of the probability of drawing a single one? Indeed, when one reflects on the matter, it is really quite curious that investigators, both here and in the "conservatism" and "gambler's fallacy" cases, should have preferred to impute a mathematical fallacy to many of their subjects rather than just to impute an implicit element of doubt whether the experimental apparatus constituted a perfect game of chance. After all, even a very long series of tests could establish no better than a high probability that it was actually a game of chance.

Perhaps subjects' protocols could usefully be sought in such experiments in order to discover whether they confirm the presence of doubt about its being a game of pure chance. It would certainly be interesting to have such protocols. But no protocols could confirm the view that laymen are inherently prone to overestimate the probabilities of conjunctions or underestimate those of disjunctions, since it is quite clear that when the task is unambiguously presented lay subjects are capable of responding in accordance with correct mathematical principles in regard to the Pascalian probability both of a conjunction (Beach, 1966) and of a disjunction (Beach and Peterson, 1966).

VI THE SIGNIFICANCE OF SAMPLE-TO-POPULATION RATIO

Another issue that deserves examination here concerns proportionality in sample size. It seems well established (Bar-Hillel, 1979) that under certain

circumstances subjects attach evidential significance to the ratio of the size of a sample to the size of its population, irrespective of the absolute size of the sample. For example, if 1000 voters have been polled in each of two cities, one with a population of 50,000 voters and the other with a population of 1,000,000 voters, most subjects are more confident in the poll taken in the smaller city. Investigators hold (e.g. Bar-Hillel, 1979, p. 251) that in such a case only absolute sample-size, and not sample-to-population ratio, is related to expected accuracy. So they think that most subjects are mistaken in the importance they attach here to proportionality.

But these subjects need not be mistaken at all. According to the Norm Extraction Method we should choose an interpretation of their responses that fits the recognisable complexity of their problem. First, we should reject any idea that they are trying to evaluate by reference to sample-to-population ratio the very same function (called by Neyman (1937) the co-efficient of confidence) that is correctly evaluated by reference to absolute sample-size. We may then suppose that these subjects are treating the sample-to-population ratio as a straightforward measure of the weight of the evidence, in the sense that, because it measures how big a fraction the actually available evidence (i.e. the actual sample) is of the total possible evidence (i.e. the population), it thereby measures how small a risk is being taken that the sample is biased in some way by its method of selection. So here we have yet another situation in which probability conceived as natural propensity or appropriate intensity of belief might differ from probability conceived in terms of proportions of possible outcomes. The proportion of samples that resemble their population in respect of some particular statistical magnitude (within a given interval of approximation) is determined just by the absolute size of the sample. But the strength with which we may legitimately believe the selected sample to resemble the population must depend also on the degree of care with which we have sought unbiased evidence from the population. One familiar method—the Baconian one—of exercising such care is to select several samples, under appropriately controlled variations of potentially relevant circumstances. Another method is to increase the fraction of the population that is sampled. The latter method is cruder than the former, and less effective if the population is very large. But it is nevertheless a step in the right direction.

Indeed, it is clear that the absolute size of the sample, on the one side, and the sample-to-population ratio, on the other, may be thought of as determining the values of two distinct functions—confidence and weight, respectively—for measuring the reliability of what the sample indicates. Subjects may therefore be expected to respond in different ways, according to whether they operate implicitly with both functions or just with one. In the former case they can take account of both weight and confidence by amalgamating them in a measure of their strength of belief within a Pascalian

framework. And in the latter case they can judge either by weight alone, in which case they will be influenced by sample-to-population ratios irrespective of absolute sample-size, or by confidence alone, in which case they will be influenced by absolute sample-size irrespective of sample-to-population ratios. According to the Norm Extraction Method experiments may thus show how many, or what kind of, people respond in one way or the other to different kinds of instructions. They cannot show that most subjects respond mistakenly unless most individual subjects give irredeemably self-contradictory responses. And such responses turn out to be very rare indeed.

VII

THE EXPERIMENTER AS EDUCATIONAL MIDWIFE

Some investigators (e.g. Pitz, 1980, p. 91) have been concerned that experimental results in this field may sometimes be artifacts of task description rather than true indications of judgmental capacity. But such a concern arises out of using the Preconceived Norm Method rather than the Norm Extraction Method. It assumes a norm to which it supposes that the subjects ought to be conforming, and points to the risk that a misleading task-description may cause deviations. The Norm Extraction Method, on the other hand, treats every response as potentially indicative of the norm that the untutored subject is cued to apply by the task-description and its circumstances, so that no such response is more "artificial" than any other.⁶

Moreover, since the Norm Extraction Method encourages investigators to think of themselves as enquiring which norms are applied by subjects in particular kinds of circumstances, it also promotes the study of which circumstances will tend to elicit the application of norms that are already known not to be applied as much as they might be. Perhaps some investigators have been too eager to devise circumstances in which subjects do not apply some computationally useful norm, such as the law that governs the significance of sample-size, and insufficiently eager to devise circumstances in which subjects' responses come to show recognition of that norm.

Results were achieved by Tversky and Kahneman (1971 and 1974) which undoubtedly demonstrate that in certain circumstances most subjects ignore the significance of sample size. For example, a set of subjects were asked to consider the following question:

A certain town is served by two hospitals. In the larger hospital about 45 babies are born each day, and in the smaller hospital about 15 babies are born each day. As you know about 50% of all babies are boys. However the exact percentage varies from day to day. Sometimes it may be higher than 50%, sometimes lower. For a period of 1 year, each hospital recorded the days on which more than 60% of the babies born were boys. Which hospital do you think recorded more such days?

Most of the subjects judged the probability of obtaining more than 60% boys to be the same in the small and in the large hospital, and Tversky and Kahneman interpret this result as being contrary to sampling theory and as indicating lay ignorance of the law of large numbers. Now, it is in any case arguable, on the Norm Extraction Method, that in these circumstances most subjects are not so much failing in an attempt to measure Pascalian probabilities correctly, as Tversky and Kahneman supposed, as succeeding in an attempt to measure Baconian ones (Cohen, 1979). They are putting a crude, causal gloss on the question and taking it to ask whether the size of a hospital exercises any causal influence on variations in the boy-girl birth ratio. But, whether or not that reinterpretation is accepted, there is no doubt that the tendency to ignore sample-size is not due to lay subjects' being inherently incapable of recognising its significance unless explicitly taught about it. Though Jones and Harris (forthcoming) replicated Tversky and Kahneman's results both in regard to the same task and in regard to a structurally similar one, they also showed that a substantial majority of subjects did appreciate the significance of sample-size in their performance of the Tversky and Kahneman task if these subjects had been tested previously on a rather simple Piagetian task which was designed to elicit awareness of this significance in a practical context. (First 5, and then 50, marbles were dropped haphazardly into a divided box, and in each case subjects were asked for the likely distribution of dropped marbles between the two parts of the box, shown the actual distribution, and asked to explain it.)

In other words subjects can be led to acknowledge and apply the law of large numbers about sample-size by a procedure which suggests that the law was already implicit in their competence as a piece of tacit knowledge. It required a mathematical genius like Bernoulli to formulate this implication of Pascalian probability for the first time. But, now that has happened, anyone who learns of the law can teach it to others, either by explicit instruction or by a maieutic method such as that of the Piagetian box. The latter method is like the way in which Socrates in Plato's *Meno* brought the mathematically uneducated young boy to a self-conscious awareness of his knowledge that, for any given square, a square on its diagonal is twice its area. This is not a way of acquiring new knowledge *from* experience, but of becoming conscious, at least for a while, of existing knowledge *through* experience. And Jones and Harris's finding suggests that investigators in the field of cognitive rationality might make some useful contributions to the technology of education if they directed their attention more to searching for circumstances that will bring people to apply correctly a particular principle that is already implicit in their cognitive apparatus than to searching for circumstances in which people do not recognise the relevance of the principle.

There is an analogy here with ambiguous drawings like Wittgenstein's

duck-rabbit. One time you see such a drawing as a duck, another time as a rabbit, and some people may not be able to see it as either unless appropriately prompted. So too the problem about the boy-girl birth ratios in the two hospitals is subject to a similar gestalt-switch. It may be seen first as a problem about causal connections and Baconian probability, and then later as one about chance distributions and Pascalian relative frequency. As pointed out in Cohen (1979, p. 401) it is pertinent that Piaget and Inhelder (1975) found that the idea of chance developed later in children than the idea of causality.⁷ Again, the Phillips and Edwards (1966) problem about predominant chip-colour—see section III above—may be seen first as a problem about Pascalian degree of justified belief, and then, after calibration (Messick and Campos, 1972), as one about *a priori* Pascalian ratios. Perhaps the more sophisticated—or better quantified—conception of such a problem, if given an opening, will tend to prevail over the cruder and less precise one. Perhaps, that is, subjects cued into the former conception are more inclined to find fault with answers given under the latter conception than subjects cued into the latter are inclined to find fault with answers given under the former. But though lay subjects' construals of their tasks may sometimes be unsophisticated, their responses are not therefore incorrect or irrational. Indeed, it would scarcely be a surprising discovery if here and there laymen were found not to achieve unaided anticipation of problems, and techniques for solving them, that form the standard subject-matter of elementary courses in statistical theory. Why else are such courses given?

Moreover there are some cases in which the transition to a more sophisticated or statistical conception of probability would not bring any worthwhile benefit. It is not at all obvious for example (see section IV above), that a careful gambler would always gain by abandoning his "fallacy". And it is certain (Cohen, 1981a, pp. 328f. and 365f.) that the taxi-cab problem is one to which in real life the statistical answer would generally be less appropriate. In this problem subjects were told that in a certain town blue and green cabs operate in a ratio of 85 to 15, respectively, while a witness identifies a cab in a crash as green and the court is told that in the relevant light conditions he can distinguish blue cabs from green ones in 80% of cases. Subjects were then asked: what is the probability (expressed as a percentage) that the cab involved in the accident was blue? If they view this as a question about degree of justified belief subjects may be expected to bear in mind also any other relevant knowledge that they may happen to have. For example, they may well be aware, quite independently, of the possibility that even a less numerous cab-fleet may dominate in a particular area of a town, or because of bad maintenance or poorly trained drivers be particularly liable to accidents, etc. With such facts at the back of their minds, subjects could reasonably regard the relative size of the two cab-fleets

as too light-weight a factor to be allowed to modify the strength of a belief that derives from fairly reliable eye-witness testimony. They would then produce what is in fact the prevalent response: the cab in the accident has a 20% probability of being blue. But if anyone views the question as being about a relative frequency—specifically, the relative frequency of correct identifications of a cab-colour in the town as blue—the correct response would be quite different: approximately 59%. This is the only response that Kahneman and Tversky (1972), Lyon and Slovic (1976) and Bar Hillel (1980) take to be correct. But its correctness depends on imposing an interpretation on the question that excludes any background knowledge of a general kind from being relevant. And in a comparable real-life situation such exclusion would not normally be either practicable or desirable, any more than it profits gamblers to forget that coins or dice are sometimes biased and decks or wheels are sometimes rigged.

In outline, therefore, the position so far revealed by the Norm Extraction Method seems to be the following, though much further experimental work no doubt remains to be done. If causal issues are in the offing, lay subjects tend to construe their problem in Baconian terms (Cohen, 1979, p. 397) and to think of probability non-complementationally as running up from non-proof to proof. This tendency is overridden in favour of Pascalian reasoning if subjects are prompted (e.g. by being asked to rate a probability by a betting-quotient or ratio) to think of probability complementationally as running up from disproof to proof. In their Pascalian reasoning subjects most naturally tend to think of probability as a propensity or as a measure of justified belief and to construe their tasks accordingly. But this tendency is overridden in its turn if subjects are prompted (e.g. by an exercise in calibration or by professional training) to think of probability in statistical terms as an *a priori* given ratio of chances or an empirically estimatable relative frequency. And in no case (see also Cohen, 1981a, p. 325) does there seem any good reason to hypothesise that subjects use an intrinsically fallacious heuristic.⁸

L. Jonathan Cohen
The Queen's College
Oxford, OX1 4AW,
U.K.

NOTES

¹ The distinction was drawn in explicit terms in Cohen (1977) p. 6, but was overlooked in Cohen (1981a) pp. 319f. and 363.

² Strictly speaking, if all the relevant knowledge we have about Jones is that he is a university teacher, then on the relative frequency analysis we estimate the probability of a person's surviving to age 70, given that he is a university teacher, while on the

personalist account we estimate the probability of a person's surviving to age 70, given the evidence about the relative frequency with which university teachers survive to age 70.

³ The connection between the two analyses is discussed in Mellor (1982).

⁴ Cf Adler (forthcoming) on how standing principles of conversational understanding may influence subjects to select an understanding of their task that involves an element of empirical judgment, rather than just logical or mathematical computation. The relevance of such principles to the interpretation of subjects' responses was stressed by Cohen (1981a, p. 326) and is now acknowledged by Kahneman and Tversky (1982, pp. 132–135).

⁵ Other interesting evidence against the prevalence of a representativeness heuristic is discussed by Olson (1976) and Evans and Dusoir (1977). But there is nothing necessarily wrong with a representativeness heuristic so far as this is just a procedure for judging the Baconian probability of a situation's outcome in terms of the situation's similarity in causally relevant respects to situations that are known to have had similar outcomes. That is the right way to assess a Baconian probability (Cohen, 1977), though of course mistakes can be made in practice if irrelevant similarities are taken to be relevant or relevant ones to be irrelevant.

⁶ I do not wish to suggest, however, that the only type of cueing to which people's judgments of probability are susceptible is in regard to relevant norms. They can be influenced not only in regard to the nature of the function with which they implicitly operate but also in several other ways, such as in regard to the domain to which they apply this function. Consider, for example, the problem discussed by Weaver (1963). The statement of the problem asserts that there are three chests, each containing two drawers, and that one chest has a gold piece in each drawer, another a silver piece in each drawer, and the third a gold piece in one drawer and a silver piece in the other. The instruction is then: "Choose a chest at random, choose one of its drawers at random, and open the drawer. Suppose it contains a gold piece. What is the probability that the other drawer in this same chest also contains a gold piece?" People tend to reply " $\frac{1}{2}$ ", whereas the propounders of the problem normally insist that the correct answer is " $\frac{2}{3}$ ". But in fact there are two different probabilities that might be at issue here. One is the probability that a randomly selected drawer contains a gold coin, given that the drawer adjacent to it contains a gold coin, and this probability is certainly $\frac{2}{3}$. The other probability is that a randomly selected chest contains two gold coins, given that it contains at least one gold coin, and this probability is equally certainly $\frac{1}{2}$. The ordinary reader or hearer of the problem is naturally cued to select the latter interpretation by having his attention directed initially towards the selection of a chest (rather than a drawer).

⁷ In their earlier papers arguing adult ignorance of the law of large numbers (Kahneman and Tversky, 1974; Tversky and Kahneman, 1971 and 1974) Tversky and Kahneman paid no attention to Piaget's evidence for children's tacit knowledge of that law. But they now acknowledge (Kahneman and Tversky, 1982, p. 130–131) the results that may be achieved by the Socratic method in this connection.

⁸ I am grateful for helpful comments on earlier drafts to Jonathan Adler, Ned Block, Radu Bogdan, Gillian Cohen, Dan Dennett, and Sir Richard Eggleston, and to participants in discussions at Cambridge, Houston, Lancaster and Tulane Universities and in Boston. I am also grateful to the British Academy for a research readership during tenure of which this paper was written.

REFERENCES

- ADLER, J. E. (forthcoming). Abstraction is uncooperative.
- ALKER, H. A. and HERMAN, M. G. 1971. Are Bayesian decisions artificially intelligent? The effect of task and personality on conservatism in processing information. *Journal of Personality and Social Psychology* **19**, 31–41.
- BAR-HILLEL, M. 1973. On the Subjective Probability of Compound Events. *Organizational Behavior and Human Performance* **9**, 396–406.
- BAR-HILLEL, M. 1979. The Role of Sample Size in Sample Evaluation. *Organizational Behavior and Human Performance* **24**, 245–257.
- BAR-HILLEL, M. 1980. The base-rate fallacy in probability judgments. *Acta Psychologica* **44**, 211–33.
- BEACH, L. R. 1966. Accuracy and Consistency in the Revision of Subjective Probabilities. *IEEE Transactions in Human Factors in Electronics* HFE-7, 29–37.
- BEACH, L. R. and PETERSON, C. R. 1966. Subjective Probabilities for Unions of Events. *Psychonomic Science* **5**, 307–308.
- CARROLL, J. S. 1980. Analysing Decision Behavior: the Magician's Audience. In T. S. Wallsten (ed.), *Cognitive Processes in Choice and Decision Behavior*, Hillsdale, New Jersey: Erlbaum, 69–76.
- COHEN, J. and HANSEL, C. M. 1957. The nature of decisions in gambling: Equivalence of single and compound subjective probabilities. *Acta Psychologica* **13**, 357–370.
- COHEN, L. J. 1977. *The Probable and the Provable*. Oxford: Clarendon Press.
- COHEN, L. J. 1980. Some historical remarks on the Baconian conception of probability. *Journal of the History of Ideas* **41**, 219–281.
- COHEN, L. J., 1979. On the psychology of prediction: whose is the fallacy? *Cognition* **7**, 385–407.
- COHEN, L. J. 1981a. Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences* **4**, 317–370.
- COHEN, L. J. 1981b. Subjective Probability and the Paradox of the Gatecrasher. *Arizona State Law Journal*, 1981, 627–34.
- DE FINETTI, B. 1931. Sul significato suggestivo del probabilità. *Fundamenta Mathematica* **17**, 298–329.
- DONMELL, M. L. and DU CHARME, M. W. 1975. The effect of Bayesian feedback on learning in an odds estimation task. *Organisational Behavior and Human Performance* **14**, 305–313.
- DU CHARME, W. M. and PETERSON, C. R. 1968. Intuitive Inference about Normally Distributed Populations. *Journal of Experimental Psychology* **78**, 269–275.
- EBBESEN, E. B. and KONEČNI, V. J. 1980. On the External Validity of Decision-making Research: What Do we Know about Decisions in the Real World? In T. S. Wallsten (ed.), *Cognitive Processes in Choice and Decision Behavior*, Hillsdale, New Jersey: Erlbaum. 21–45.

- EVANS, J. St. B. T. and DUSOIR, A. E. 1977. Proportionality and sample size as factors in intuitive statistical judgment. *Acta Psychologica* **41**, 129–137.
- HACKING, I. 1975. *The Emergence of Probability*. Cambridge: Cambridge U.P.
- JONES, C. J. and HARRIS, P. L. (forthcoming) Insight into the Law of Large Numbers: a Comparison of Piagetian and Judgement Theory. *Quarterly Journal of Experimental Psychology* (in press).
- KAHNEMAN, D. and TVERSKY, A. 1972. On the Psychology of Prediction. *Oregon Research Institute Research Bulletin* **12**, whole No. 4.
- KAHNEMAN, D. and TVERSKY, A. 1973. On the Psychology of Prediction. *Psychological Review* **80**, 237–251.
- KAHNEMAN, D. and TVERSKY, A. 1974. Subjective Probability: A Judgment of Representativeness. In C.-A.S. Staël von Holstein (ed.), *The Concept of Probability in Psychological Experiments*. Dordrecht: Reidel.
- KAHNEMAN, D. and TVERSKY, A. 1982. On the Study of Statistical Intuitions. *Cognition* **11**, 123–141.
- KAHNEMAN, D. and TVERSKY, A. (forthcoming) Intuitive Predictions: Biases and Corrective Procedures. *Management Science* (in press).
- KYBURG, H. E., JR. and SMOKLER, H. E. 1964. Introduction. In H. E. Kyburg and H. E. Smokler (eds.), *Studies in Subjective Probability*. New York: John Wiley.
- LAKATOS, I. 1970. "Falsificationism and the Methodology of Scientific Research Programmes", in *Criticism and the Growth of Knowledge*, ed. I. Lakatos and A. E. Musgrave, Cambridge: University Press, 91–195.
- LINDMAN, H. and EDWARDS, W. 1961. Unlearning the Gambler's Fallacy. *Journal of Experimental Psychology* **62**, 630.
- LYON, D. and SLOVIC, P. 1976. Dominance of accuracy information and neglect of base rates in probability estimation. *Acta Psychologica* **40**, 287–98.
- MACKIE, J. L. 1973. *Truth, Probability and Paradox*. Oxford: Clarendon Press.
- MELLOR, D. H. 1982. Chance and Degrees of Belief. In R. McLAUGHLIN (ed.), *What? Where? When? Why: Essays on Induction, Space and Time, Explanation*, Dordrecht: Reidel, 49–68.
- MESSICK, D. M. and CAMPOS, F. T. 1972. Training and Conservatism in Subjective Probability. *Journal of Experimental Psychology* **94**, **3**, 335–337.
- NAGEL, E. 1955. Principles of the Theory of Probability. In O. Neurath, R. Carnap, C. Morris (ed.), *Foundations of the Unity of Science*, Chicago: Chicago U.P., Vol. I, 341–422.
- NEYMAN, J. 1937. Outline of a Theory of Statistical Estimation, Based on the Classical Theory of Probability. *Philosophical Transactions of the Royal Society of London*, Series A, 236, 333–380.
- NISBETT, R. and ROSS, L. 1980. *Human Inference: strategies and shortcomings of social judgment*, Englewood Cliffs, New Jersey: Prentice-Hall.
- OLSON, C. L. 1976. Some Apparent Violations of the Representativeness

- Heuristic in Human Judgment. *Journal of Experimental Psychology: Human Perception and Performance* **2**, 599–608.
- PETERSON, C. R., ULEHLA, Z. J., MILLER, A. J., BOURNE, L. E. and STILSON, W. A. 1965. Internal Consistency of Subjective Probabilities. *Journal of Experimental Psychology* **70**, 526–533.
- PETERSON, C. R. and DU CHARME, W. M. 1967. A Primacy Effect in Subjective Probability Revision. *Journal of Experimental Psychology*, **73**, **1**, 61–65.
- PHILLIPS, L. D. and EDWARDS, W. 1966. Conservatism in a Simple Probability Inference Task. *Journal of Experimental Psychology*, **72**, **3**, 346–354.
- PIAGET, J. and INHELDER, B. 1975. *The Origin of the Idea of Chance in Children*. London: Routledge.
- PITZ, G. F., DOWNING, L. and REINHOLD, H. 1967. Sequential Effects in the Revision of Subjective Probabilities. *Canadian Journal of Psychology* **21**, 381–393.
- PITZ, G. F. 1980. The Very Guide of Life: The Use of Probabilistic Information for Making Decisions. In T. S. Wallsten (ed.), *Cognitive Processes in Choice and Decision Behavior*, Hillsdale, New Jersey: Erlbaum, 77–94.
- RAMSEY, F. P. 1931. *The Foundations of Mathematics*, London: Routledge and Kegan Paul.
- SAVAGE, L. J. 1954. *The Foundations of Statistics*, New York: Wiley.
- SAVAGE, L. J. 1964. The Foundations of Statistics Reconsidered. In H. E. Kyburg and H. E. Smokler (eds.), *Studies in Subjective Probability*, New York: John Wiley.
- SHAFER, G. 1978. Non-Additive Probabilities in the Work of Bernoulli and Lambert. *Archive for History of Exact Sciences* **19**, 309–370.
- SLOVIC, P., FISCHHOFF, B., and LICHTENSTEIN, S. 1977. Behavioral Decision Theory. *Annual Review of Psychology* **28**, 1–39.
- SWINBURNE, R. G. 1964. "Falsifiability of Scientific Theories", *Mind* **73**, 434–436.
- TVERSKY, A. and KAHNEMAN, D. 1971. The belief in the law of small numbers. *Psychological Bulletin* **76**, 105–110.
- TVERSKY, A. and KAHNEMAN, D. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science* **185**, 1124–1131.
- VLEK, C. A. J. and VAN DER HEIJDEN, A. H. C. 1967. Subjective likelihood functions and variations in the accuracy of probabilistic information processing. *Psychological Institute Report* no. E 017-67, University of Leyden.
- VLEK, C. A. J. 1973. Coherence of Human Judgment in a Limited Probabilistic Environment. *Organizational Behavior and Human Performance* **9**, 460–481.
- WATKINS, J. W. N. 1958. Confirmable and Influential Metaphysics, *Mind* **67**, 344–365.
- WEAVER, W. 1963. *Lady Luck: the Theory of Probability*. London: Heinemann.