Brit. J. Phil. Sci. **36** (1985), 331–340 *Printed in Great Britain*

Review Articles

ILLUSIONS ABOUT UNCERTAINTY*

This collection of essays by leading researchers in the way humans form judgements of uncertainty and risk is dominated by an interest in the role of 'heuristics' in the assessment of uncertainty and, in particular, with the 'biases' which the use of such heuristics can induce in these assessments.

The basic theme is philosophically and psychologically intriguing. Mathematicians and statisticians often use simplifications in calculations which give results sufficiently accurate for the task at hand. Such expedients are often more than mere computational conveniences. They become necessities when we lack the capacity to derive exact solutions even when hooked up to the latest technology. The use of such techniques of approximation in scientific work is accompanied, of course, by serious efforts to determine the conditions when the approximations are good or bad. Both poisson and normal distributions can approximate binomial distributions; but they perform well under different circumstances. It is important to keep control over the application of such techniques to avoid excessive bias in estimates made with their aid.

I have been alluding to techniques for cutting corners in calculation which are deliberately instituted to provide good if not exact estimates where exact estimates are too costly or not feasible at all.

It is entertainable, however, that the intuitive reasoner cuts corners in his deliberations using techniques acquired by nature or nurture in a manner beyond the control of the reasoner or those involved in his upbringing. And one might conjecture that such techniques are acquired to deal with tasks whose exact solutions would tax the abilities of the intuitive reasoner. If that is the case, it is to be expected that these techniques should enable the intuitive reasoner to perform well at tasks of computation and ratiocination ordinarily demanded of her or him especially if survival value were a contributory explanatory factor in accounting for the acquisition of the rule.

On the other hand, it is also to be expected that sometimes such techniques should lead to error and, indeed, should do so even when the computational capacities needed to obtain a more accurate result are available to the agent. Because the use of the techniques are part of the agent's unreflective repertoire of reasoning practices, she or he exercises no critical control over their use and so relies on them without a moment's thought even in contexts where they go badly wrong.

* Review of Daniel Kahneman, Paul Slovic and Amos Tversky (*eds.*) [1982]: *Judgement under Uncertainty: Heuristics and Biases*. Cambridge University Press.

Kahneman and Tversky conjecture that in our intuitive and, indeed, even in our fairly deliberate estimates of probability and risk we use just such unreflectively acquired reasoning practices or 'heuristics'. They offer some characterisation of certain types of heuristics; and because the operation of such heuristics might be expected to be most clearly manifested in the performance of relatively simple tasks where even sophisticated subjects exhibit tendencies to error, they undertake to illustrate and support their conjectures by describing investigations where experimental subjects display 'cognitive illusions' akin to perceptual illusions like the Muller–Lyer effect.

The aim of the studies of Kahneman and Tversky and their colleagues should not be understood as showing how perversely irrational and stupid humans can be in performing moderately simple tasks of estimation. The demonstrations of such irrationality and stupidity they offer are intended to furnish evidence for their conjecture that both our successes and our failures at tasks of estimating uncertainty and risk are guided by the types of heuristics they describe. So understood, the results they report do not point to incapacities so debilitating as to lead us to question the acceptability of the standards for providing exact solutions to problems of probability estimation from which intuitive reasoners stray. 'Ought' undoubtedly implies 'can'. But we are not enjoined to conform to some system of canons of probabilistic reasoning (such as the Bayesian canons which Kahneman and Tversky tend to favour) but only to conform to such canons in so far as we are not prevented from doing so due to limitations of memory, computational capacity or emotional stability. If nature and nurture saddle us with heuristics for making judgements which distort our assessments, that furnishes an excuse for our failure to live up to the norms. At the same time and more importantly, the study of such heuristics can point us in the direction of techniques which enable human reasoners to transcend the biases generated by heuristics through converting such heuristics into something more closely akin to deliberately instituted methods of making approximations which we employ on occasions when they are sufficiently accurate and avoid when they serve us poorly.

Of course, the notion of a heuristic that induces bias needs fleshing out in order for it to qualify as a conjecture amenable to serious critical scrutiny, empirical test and possible acceptance. Participants in this anthology undertake efforts to elaborate these conjectures. Heuristics like representativeness and availability have a prominent place in the discussion; but to the reader who is not a participant in the research, a grasp of the character of such heuristics is obtained chiefly from illustrations of the heuristics in action as brought out by the interesting experimental studies reported in this volume.

The ideas which emerge are sometimes suggestive and surely worthy of being explored further. But the efforts to show how the heuristic mechanisms operate in the explanation of the experimental data remain vague and

elusive. They do not qualify as yet as explanatory hypotheses which can clarify hithertofore dark corners of our knowledge.

As an illustration consider the following alleged example of a bias induced through the inappropriate operation of the so called 'representativeness' heuristic:

> All families of six children in a city were surveyed. In 72 families the *exact order* of births of boys and girls was GBGBBG. What is your estimate of the number of families surveyed in which the *exact order* of births was BGBBBB?

Kahneman and Tversky report that most respondents judge the second sequence less likely than the first (p. 34) even though the probabilities are objectively roughly the same. This is explained by the fact that the relative frequency of boys and girls in the first sequence is regarded as more representative of the proportion of boys and girls in the total population than is the proportion in the second sequence.

Examples such as this are alleged to support the conjecture that a heuristic (in this case representativeness) is operating because errors are made even though there is no great obstacle due to computational complexity preventing the subjects from coming to the correct answer. Given the relatively widely understood fact that the probabilities of a boy and of a girl being born are roughly equal, the conclusion that the two sequences are equally likely or nearly so is readily drawn. Yet the experimental subjects overlook this point and leap to a mistaken conclusion due, so claim Kahneman and Tversky, to the operation of the representativeness heuristic.

It may, perhaps, be objected that the experimental subjects need not be understood to have made errors in reasoning but merely to have ignored the information about the order of births. Such inattention to instructions is, to be sure, an error; but it is not an error in reasoning about probabilities. Kahneman and Tversky respond to this observation by reporting the results of comparisons made by experimental subjects of the first sequence GBGBBG with the sequence BBBGGG. The experimental subjects regarded the latter less likely than the former. Here the two sequences exhibit the same frequencies of boys and girls but differ with respect to order. Kahneman and Tversky suggest that this might be attributed to the fact that the sequence BBBGGG exhibits a less random order than the sequence GBGBBG. They argue that experimental subjects do not neglect order entirely so that the first mistake cannot be blamed on the experimental subjects' neglect of order.

This conclusion, in my judgement, is too hasty. Perhaps the aspect of order taken into account is the randomness of the sequence. GBGBBG exhibits a more random order than BGBBBB just as it displays greater randomness than BBBGGG. If the experimental subject is taking the randomness of the order into account in making these comparisons (in spite

of the explicit instructions to take the exact order of the sequence into account), there is no error in his reasoning but, once more, a failure on his part to follow the instructions. The greater the randomness in the order, the more likely it is that a sequence of that type will occur.

To be sure, the experimental subject has made a mistake in reading the instructions. The comparisons with respect to probability were explicitly supposed to be for sequence in the specific orders given. The experimental subjects ignore the question as explicitly asked and substitute another. No doubt they make mistakes; but they do not commit errors in reasoning.

I do not claim that my interpretation of the results reported by Kahneman and Tversky is sound. To decide between my construal and theirs requires more experimental investigation. My point is rather that it is not easy to understand or assess the merits of the conjectures put forth by Kahneman and Tversky from the experimental results they report.

Two problems emerge from examining this particular example which resemble difficulties with other topics reviewed in this volume:

(1) Kahneman and Tversky understood fully that representativeness may be rendered precise only by specifying the features of the situation with respect to which representativeness is guiding the experimental subject. But they fail to appreciate how this fact threatens to trivialise their thesis concerning representativeness. I can explain both comparisons of sequences with respect to probability as relying on representativeness with respect to randomness of the orderings. Kahneman and Tversky use representativeness with respect to frequency in one comparison and representativeness with respect to randomness of order in another. Surely the explanatory force of their conjecture does not derive from the fact that representativeness with respect to something or other is being invoked by experimental subjects. The explanatory force of an appeal to the representativeness heuristic ought to invoke a notion of representativeness with respect to something fairly specific. Otherwise the appeal to the representativeness heuristic does not tell us very much.

(2) Assuming we have a clear conception of the respect in which something is representative of something else we are assessing, we must still ask whether the role of the representativeness heuristic is to direct the way the experimental subject interprets the task being set or to control the manner in which the experimental subject reasons on the information given to her or him. When speaking in a general way, Kahneman and Tversky suggest that heuristics guide probabilistic reasoning and not the interpretation of the task being set. But in the example just described and many others, the error could very well be in the representation and interpretation of the problem and not in the reasoning given an interpretation of the task.

The issue is of some importance. If the heuristics are intended to be procedures for making estimates or other evaluations of probabilities which

typically yield outputs in agreement with correct reason but sometimes do not, the errors exhibited in the examples will not be as striking confirmations of the presence of the conjectured heuristics as they appear to be if they are errors in the interpretation of tasks demanded by the questions asked rather than errors in the reasoning. To be sure, it might be held that if one is wedded to a certain heuristic, one will exhibit a tendency to reconstrue as much as one can so as to allow the heuristic to operate; but if we do adopt this response, we come perilously close to losing whatever predictive and explanatory power the conjectures under consideration seem to have.

I want to emphasise that I am not committed to defending the virtue of experimental subjects as expert reasoners. I am worried rather about the extent to which experiments of the sort reported by Kahneman and Tversky and their colleagues really do support their interesting conjectures about heuristics and their role in assessing risk and uncertainty.

Although I applaud the refusal on the part of participants in this volume to conclude from the fact that experimental subjects in particular and human beings in general deviate from given norms in striking ways that the norms are inadequate as ideals of rationality, there is an unfortunate tendency on their part to interpret judgements of uncertainty as rational if they meet the requirements of the calculus of probabilities. Indeed, many authors represented in the volume, including Kahneman and Tversky in particular, appear to think of Bayesian norms as paradigmatic benchmarks from which to assess deviations from norms of rationality.

For those who, like myself, think that Bayesian norms have limited applicability as prescriptive norms not because human agents cannot meet the requirements imposed but because the requirements are inappropriate expressions of ideals of rationality, this bias is more unfortunate than the biases they report in experimental subjects.

In all fairness, however, it should be said that worries of this sort are acknowledged in the final chapter. There Kahneman and Tversky concede that judgements of uncertainty may not always be probabilistic. Brief reference is made to the ideas of L. J. Cohen and G. Shafer who espouse 'non complementary' measures of degree of belief. (A complementary measure requires for every H and G that the degree of belief that H plus the degree of belief that not $-H$ equal the degree of belief that Gf plus the degree of belief that $-G$ (p. 520).)

It is to be hoped that these passing remarks will become the occasion for empirical research by Kahneman and Tversky and their colleagues in the future. In this anthology, they appear as an afterthought. The literature on non probabilistic measures of uncertainty which developed in modern times starting with G. L. S. Shackle is essentially neglected. There is little appreciation of the distinction between views like Shackle's or Cohen's which seek to identify intended applications of the measures proposed and views like Shafer's which confuse Shackle's ideas with lower probablities in

the rather special understanding of lower probability proposed by A. P. Dempster in the 1960s.[1]

Worse yet, Kahneman and Tversky omit from consideration notions of grades of support, confirmation, likelihood or belief (terminology is by no means standard) which are neither expectation determining probabilities of the sort aptly represented by measures satisfying the calculus of probabilities nor Shackle-like measures of belief (or degrees of disbelief which Shackle called degrees of potential surprise).

The distinctions just made are of some relevance to the topics discussed in this anthology and should, perhaps, be developed further. There are at least three different types of tasks addressed by assessments of uncertainty, degrees of probability and the like. The way such assessments are understood by agents ought not be the same relative to all of these tasks.

One such task is the assessment of probabilities to be used in calculating expected utilities or values of policies in decision making. I shall take for granted that probabilities meeting the requirements of at least a finitely additive probability are appropriate for this purpose.

Sometimes it is alleged that when deciding which of a roster of potential answers to a given question to accept, one should accept answers whose warrant or support is high enough or great enough. If 'probability' is used for this type of assessment, it may be called 'satisficing probability'.

Sometimes in deciding what to accept, we speak of maximising support or probability rather than satisficing it.

Authors bedazzled by probabilities meeting the formal requirements of the calculus of probabilities as applied to the determination of expectations have thought that such expectation determining probabilities may also be used as satisficing or maximising probabilities as well.

But if expectation determining probabilities are identical with both satisficing and maximising probabilities, no one may accept a hypothesis unless it has a probability 1 relative to the available background knowledge and evidence. Dogmatic Bayesians may be happy with this result but it surely does violence to many principled philosophical views and some strands of presystematic precedent.

If we reject the equation, we must say either that high expectation determining probability is not sufficient for acceptance or we ought not to maximise expectation determining probability.

Popper is no friend of acceptance in any sense in which acceptance involves nondeductively adding new information to knowledge already available. He is prepared to accept hypotheses, however, as conjectures for the purpose of test and development. He pointed out a long time ago that if we seek to maximize expectation determining probability, we shall accept nothing unless it is entailed by the evidence and background knowledge.

[1] See I. Levi [1984]: *Revisions and Decisions*, ch. 14. Cambridge University Press.

This point holds whether we use his sense of 'acceptance' or use a sense which admits the viability of adding new information to knowledge which does not entail it (*i.e.* admits the legitimacy of nonductive or inductive inference). Both the anti-inductivist Popper and unreconstructed inductivists like myself can agree relative to our different conceptions of acceptance that there is something wrong with the idea of maximising expectation determining probability.

Nor is high expectation determining probability sufficient for acceptance—at least if we seek a notion of acceptance which preserves the deductive closure of what is accepted (and still avoids 'lottery paradoxes').

These considerations suggest that there are three different types of task for which we concoct notions of support or degrees of warranted belief and, perhaps, degrees of unwarranted belief. They are not to be explicated in the same way; but it is at least entertainable that they all surface in some contexts or other where psychologists ask experimental subjects to assess likelihoods or probabilities (since experimental subjects may not be supposed to be sticklers for the niceties of technical talk unless the experimental design has been controlled to secure the acceptability of that assumption.

I have long favoured the view (and L. J. Cohen follows me in this) that Shackle-like measures are promising characterisations of the formal structure of measures of degrees of belief which are satisficed in acceptance.

When it comes to indices to be maximised, several types of measures have been proposed. Likelihoods (probabilities of experimental outcomes given statistical hypotheses), measures of relevance such as differences or ratios of posterior and prior probabilities or differences in their logarithms, Popperian measures of corroboration, Good's measures of explicativity and measures of expected epistemic utility all purport to be indices to be maximised in comparing hypotheses as better or worse on the evidence.

Failure to attend to these distinctions has awkward implications for some widely cited results reported by Kahneman, Tversky and their colleagues. Thus, this anthology contains references to a so called 'conjunction effect' used to illustrate the workings of the representativeness heuristic.

Experimental subjects are invited to rank a series of propositions with respect to probability including propositions A, B and $A\&B$. They are to base their judgements on information furnished by a character sketch which describes a given individual as representative of persons in one category and unrepresentative of persons in another. A asserts that the individual is in the first category. B asserts that the individual is in the second. $A\&B$ asserts that the individual is in both categories. Experimental subjects do tend to rank the three propositions with respect to representativeness of the information in the character sketch so that A is most representative, $A\&B$ is next and B is least representative. When asked to assess the three propositions with respect to probability, they rank the propositions in exactly the same way. This last ranking, however, violates the requirements of the calculus of probabilities.

As an example consider the following character sketch used by Kahneman and Tversky in some of their studies (p. 92):

> Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

Relative to this information, experimental subjects were asked to rank several propositions with respect to probability including the following three:

Linda is active in the feminist movement. (A)
Linda is a bank teller. (B)
Linda is a bank teller and active in the feminist movement. ($A\&B$)

Both sophisticated and unsophisticated subjects in overwhelming numbers are reported to exhibit the conjunction effect—*i.e.* to regard $A\&B$ as more probable than B. The same ranking is also exhibited when they are asked to rank the propositions with respect to representativeness.

Examples such as these are taken to illustrate deviations from the calculus of probabilities due to the impact of the representativeness heuristic. If the experimental subjects do indeed understand the task they face as requiring the assessment of expectation determining probabilities, then their responses do deviate from the requirements of probability theory. But when someone is asked to rank propositions relative to data, it is quite easy for that person to construe his or her task to be to determine the grades of acceptability of the propositions on the data—*i.e.* to determine maximising probabilities. It is well known that evidence often supports (in the maximising sense) stronger hypotheses better than weaker ones. In spite of the differences between them, Popperians, those Bayesians who explicate support in terms of measures of relevance and those who, like myself, use measures of expected epistemic utilities as the indices to be maximised in deciding what to accept can agree on this point. If experimental subjects are assessing maximising probability as they might naturally be expected to be in the Linda problem, they commit no conjunction fallacy.

In work not reported in this anthology, Kahneman and Tversky describe other experiments where a type of conjunction fallacy is committed and where it is not at all plausible to suppose that the experimental subjects are assessing maximising probability.[1] These results lend some credibility to the conjecture that the same factors which induce experimental subjects to make errors in these cases induce them to rank the propositions as they do in problems like the 'Linda' example.

But even if this point is granted, it still remains the case that in the Linda

[1] A. Tversky and D. Kahneman [1983]: 'Extensional *vs.* Intuitive Reasoning: The Conjunction Fallacy in Probability Judgement', *Psychological Review*, pp. 293–315.

example and cases like it the experimental subjects may not be committing any error at all. This is important; for Kahneman and Tversky often leave the reader with the impression that errors are being made. If we are concerned to identify when natural heuristics like representativeness do and do not lead us astray, it is important to know when we are being led astray.

Furthermore, Kahneman and Tversky seem sometimes to admit that in examples like the 'Linda' problem, experimental subjects are more difficult to persuade that they have committed a fallacy than in other cases where a deviation from the requirements of the multiplication theory of the calculus of probabilities is exhibited. But they offer no explanation for this resistance to correction.

Here is a brief sketch of a conjectural explanation. Let us grant that assessments of probability are guided by representativeness in the cases where conjunctions are ranked over conjuncts in a manner which correlates well with assessments of representativeness. In some cases, however, the assessment made is not appropriate to the task which the agent takes himself to be asked to do and in other cases it is. In the former kind of case, it is easier for the experimenter to persuade the subject that he or she has made an error than in the latter kind of case.

This is, of course, only a conjecture requiring both more precise formulation and empirical investigation. It seems worth exploring because it can accommodate the considerations Kahneman and Tversky can adduce in favour of the role of the representativeness heuristic in those cases where experimental subjects turn away from the demands of the calculus of probability while insisting (rightly so in my view) that often enough though not always they ignore these demands for good reason—namely, because the task they think they are undertaking is the assessment of maximising probability and not expectation determining probability.

The treatment of the so called conjunction effect found in papers in this volume as well as elsewhere illustrates well the excessive devotion to the conception of expectation determining probability as the relevant ideal relative to which deviations from correct or valid standards of assessment of uncertainty should be studied. As I have already said, there is some indication in the final chapter that participants in this volume may be taking a broader view of this issue than the main essays in the anthology indicate. Further movement along these lines can have only a beneficial effect on future research bearing on the conjectures which Kahneman and Tversky have proposed for our consideration.

In spite of their partiality to Bayesian ideology, there are some contexts where the participants in the anthology do not get the ideology right. The best known case of this is the neglect of base rates. Kahneman, Tversky and their colleagues identify correct answers to many of the problems they set for experimental subjects and claim that the correctness of the answers is grounded on Bayesian principles when this is simply not the case. Since I have elaborated on this matter elsewhere, I shall not rehearse that discussion

here.[1] Still it is important to keep in mind that empirical studies exploring the effects of heuristic procedures to determine when they do and do not lead to error should be careful that they have a clear understanding of the normative principles which characterise when error and fallacy are present.

Finally, it seems to me that contributors to this volume have neglected the circumstance that reasonable men may legitimately refuse to make numerically definite judgements of expectation determining probability. There has been a small spurt of interest recently in experimental research into judgements of indeterminate probability and one may hope that this recent development will receive more support in the future. Unfortunately this development is not represented in this volume.

None of the reservations and objections I have advanced undermine the main ideas driving the project proposed by Kahneman and Tversky. As I suggested at the outset, there is much to sympathise with in the ideas concerning the relations between ideals of rationaility and empirical studies into cognitive performance reported on in this anthology. The criticisms I have enumerated seek to emphasise a dimension of the task Kahneman, Tversky and their colleagues have undertaken which, though recognised by them, has not received the attention it deserves.

It is crucial to the line of inquiry being promoted by Kahneman and Tversky that errors and fallacies in reasoning be investigated and their sources identified when feasible. But there is no way to understand what is to count as an error without invoking some regulative standard of ideal rationality. When experimental psychologists undertake such investigations, they ought to be sensitive to the controversies and shared agreements concerning such ideals as may exist. An excessive commitment to one side of the disputes which so often rage concerning probabilistic rationality can often depreciate the value of the experimental results obtained. The programme of research and the conception of the relation between regulative ideals of rationality and actual behaviour and reasoning which characterize the researches of Kahneman and Tversky and the other investigators represented in this volume are too interesting and important to be bounded by the parochial confines of Bayesian ideology.

ISAAC LEVI
Columbia University

[1] I. Levi [1983]: 'Who commits the base rate fallacy?' *The Behavioral and Brain Sciences*, **6**, pp. 503–6. In their response to my remarks ([1983]: *B&Bs*, pp. 509–10), Kahneman and Tversky appear to concede that Bayesians need not invoke the base rates cited in their well known taxicab problem without admitting that this undermines their own previous analyses of this and kindred problems.