The Behavioral and Brain Sciences

Editor
Stevan Harnad
P.O. Box 777, Princeton, NJ 08540

Assistant Editor
Helaine Randerson

Associate Editors
Animal Intelligence
David Premack/U. Pennsylvania

Behavioral Biology
Jack P. Hailman/U. Wisconsin

Cognition and Artificial Intelligence
Zenon Pylyshyn/U. Western Ontario

Developmental Psychology
Charles J. Brainerd/U. Western Ontario

Ethology and Neuroethology
John C. Fentress/Dalhousie

Evolutionary Biology
Michael T. Ghiselin/U. Utah

Experimental Analysis of Behavior
A. Charles Catania/U. Maryland Baltimore County

Higher CNS Function
Robert W. Doty/U. Rochester

History and Systems
Julian Jaynes/Princeton

Language and Cognition
Peter Wason/University College London

Neurobiology
Graham Hoyle/U. Oregon

Neurophysiology
Barry H. Smith/National Institutes of Health

Neuropsychology
Karl Pribram/Stanford

Philosophy
Daniel C. Dennett/Tufts

Psychobiology
David S. Olton/Johns Hopkins

Editorial Policy

THE BEHAVIORAL AND BRAIN SCIENCES (BBS) is an international journal providing a special service called Open Peer Commentary* to researchers in any area of psychology, neuroscience, behavioral biology, or cognitive science who wish to solicit, from fellow specialists within and across these BBS disciplines, multiple responses to a particularly significant and controversial piece of work. (See *Instructions for Authors and Commentators*, inside back cover.) The purpose of this service is to contribute to the communication, criticism, stimulation, and particularly the unification of research in the behavioral and brain sciences, from molecular neurobiology to artificial intelligence and the philosophy of mind.

Papers judged by the editors and referees to be appropriate for Commentary are circulated to a large number of commentators selected by the editors, referees, and author to provide substantive criticism, interpretation, elaboration, and pertinent complementary and supplementary material from a full cross-disciplinary perspective. The article, accepted commentaries, and the author's response then appear simultaneously in BBS.

The service of Open Peer Commentary is primarily devoted to original articles appearing first in BBS, but it will occasionally be extended to a recent book or already published paper widely judged to be particularly appropriate for Commentary. The editorial office will consider recommendations from the BBS readership concerning possible topics, authors or already published materials for future BBS Commentary. (Please consult the criteria for accepting manuscripts in the *Instructions*, inside back cover.)

Commentary on BBS articles may be provided by any qualified professional in the behavioral and brain sciences, but much of it is drawn from a large body of BBS Associates who have become formally affiliated with the project.

A qualified professional is eligible to become a BBS Associate if he or she has (1) been nominated by a current BBS Associate, (2) refereed for BBS, or (3) had a commentary or article accepted for publication. A special subscription rate of $29.00/£16.00 (1981) is available to Associates. Individuals interested in serving as BBS Associates are asked to write to the editor.

This publication was supported in part by NIH Grant LM 03539 from the National Library of Medicine.

*Modelled on the 'CA Comment' service of the journal *Current Anthropology*.

Copying

This journal is registered with the Copyright Clearance Center, (21 Congress St., Salem, MA 01970). Organizations in the U.S.A. who are also registered with the CCC may therefore copy material (beyond the limits permitted by sections 107 and 108 of U.S. Copyright Law) subject to payment to the CCC of the per-copy fee indicated in the code on the first page of the article. This consent does not extend to multiple copying for promotional or commercial purposes.

ISI Tear Sheet Service, 325 Chestnut Street, Philadelphia, PA 19106, is authorized to supply single copies of separate articles for private use only.

For all other use, permission should be sought from the Cambridge or New York offices of the Press.

Subscriptions

The Behavioral and Brain Sciences (ISSN 0140-525X) is published quarterly in March, June, September and December. Four parts form a volume. The subscription price, which includes postage, of Volume 4 (1981) is US $89.00 net in the U.S.A. and Canada (£48.00 in the U.K. and rest of the world) for institutions; US $44.50 net (£24.00) for individuals; and US $29.00 net for students (in the U.S.A. and Canada only) who provide proof of eligibility with order. Single parts cost US $25.00 net (£14.00) plus postage. Orders may be sent to a bookseller or, in the U.S.A. and Canada direct to: Cambridge University Press, 32 East 57 Street, New York, N.Y. 10022; in the U.K. and rest of the world to: Cambridge University Press, The Edinburgh Building, Shaftesbury Road, Cambridge CB2 2RU, England. Second class postage paid at New York, N.Y. and at additional mailing offices. Postmaster: send address changes to Cambridge University Press, 32 East 57 Street, New York, N.Y. 10022.

Advertising

Inquiries about advertising should be sent to the Journals Promotion Department of the Cambridge or New York office of Cambridge University Press.

THE BEHAVIORAL AND BRAIN SCIENCES (1981) 4, 317–370
Printed in the United States of America

Can human irrationality be experimentally demonstrated?

L. Jonathan Cohen
The Queen's College, University of Oxford, Oxford OX1 4AW, England

Abstract: The object of this paper is to show why recent research in the psychology of deductive and probabilistic reasoning does not have "bleak implications for human rationality," as has sometimes been supposed. The presence of fallacies in reasoning is evaluated by referring to normative criteria which ultimately derive their own credentials from a systematisation of the intuitions that agree with them. These normative criteria cannot be taken, as some have suggested, to constitute a part of natural science, nor can they be established by metamathematical proof. Since a theory of competence has to predict the very same intuitions, it must ascribe rationality to ordinary people.

Accordingly, psychological research on this topic falls into four categories. In the first, experimenters investigate conditions under which their subjects suffer from genuine cognitive illusions. The search for explanations of such performance errors may then generate hypotheses about the ways in which the relevant information-processing mechanisms operate. In the second category, experimenters investigate circumstances in which their subjects exhibit mathematical or scientific ignorance: these are tests of the subjects' intelligence or education. In the third and fourth categories, experimenters impute a fallacy where none exists, either because they are applying the relevant normative criteria in an inappropriate way or because the normative criteria being applied are not the appropriate ones.

Keywords: competence; deduction; fallacy; intuition; probability judgments; rationality; reasoning

Introduction

The experimental study of human rationality – that is, of validity in deductive or probabilistic reasoning – has become entangled during the past decade or so in a web of paradox. On the one hand, reputable investigators tell us that certain psychological discoveries have "bleak implications for human rationality" (Nisbett & Borgida 1975), or that "for anyone who would wish to view man as a reasonable intuitive statistician, such results are discouraging" (Kahneman & Tversky 1972b), or that "people systematically violate principles of rational decision-making when judging probabilities, making predictions, or otherwise attempting to cope with probabilistic tasks" and they "lack the correct programs for many important judgmental tasks" (Slovic, Fischhoff & Lichtenstein 1976). On the other hand, those investigators are reminded that people could not even drive automobiles unless they could assess uncertainties fairly accurately (Edwards 1975). The ordinary person is claimed to be prone to serious and systematic error in deductive reasoning, in judging probabilities, in correcting his biases, and in many other activities. Yet, from this apparently unpromising material – indeed, from the very same students who are the typical subjects of cognitive psychologists' experiments – sufficient cadres are recruited to maintain the sophisticated institutions of modern civilisation. Earlier decades, in an era of greater optimism, may well have overestimated the natural reasoning powers of human beings. But there seems now to be a risk of underestimating them.

What is needed here is a conceptual framework within which to think coherently about problems of cognitive rationality and the relevant experimental data, and the object of the present paper is to sketch such a framework. For this purpose it is necessary first of all to examine the credentials of those normative theories by reference to which investigators may legitimately evaluate the rationality or irrationality of a naive subject's inference or probability judgment. Such a normative theory, I shall argue, is itself acceptable for the purpose only so far as it accords, at crucial points, with the evidence of untutored intuition. This thesis was also argued long ago by Goodman (1954, pp. 66–67). But the argument needs to be expanded and fortified against more recent opposition. What then follows from the thesis is that ordinary human reasoning – by which I mean the reasoning of adults who have not been systematically educated in any branch of logic or probability theory – cannot be held to be faultily programmed: it sets its own standards. Of course, various kinds of mistakes are frequently made in human reasoning, both by laboratory subjects and in ordinary life. But in all such cases some malfunction of an information-processing mechanism has to be inferred, and its explanation sought. In other words, the nature of the problem constrains us to a competence-performance distinction. Our fellow humans have to be attributed a competence for reasoning validly, and this provides the backcloth against which we can study defects in their actual performance. That is the theme to be argued in the first part of the paper.

At the same time, allegations of defects in perfor-

© 1981 Cambridge University Press 0140-525X/81/030317-54/$04.00/0

mance need to be carefully scrutinised. Some of these allegations are correct and important. But others seem to arise from a misapplication or misconception of the relevant standards of rationality by which the experimentally revealed phenomena should be judged, even when those phenomena themselves are quite robust and incontestable. The second part of the paper will therefore suggest four categories to which a critical assessment of existing allegations of performance defects might appropriately assign them.

In sum, those who once tended to exaggerate human reasoning powers may be construed as having concentrated their attention too much on the facts of competence, while those who have more recently tended to underestimate these powers have concentrated their attention too much on the facts of performance, and in some cases have judged these facts too harshly.

I. The argument for rational competence

1. Intuitions as the basis of normative criteria for the evaluation of deductions. Investigators who wish to evaluate the validity of their subjects' deductions would turn naturally to some educationally well regarded textbook of formal logic, such as Quine (1952), Copi (1954) or Lemmon (1965). The assumption would be that all and only the rules of inference that are given or derivable in those systems of so-called natural deduction are valid principles of deducibility, so far as deducibility hinges on the interplay of the logical particles "not," "and," "or," "if," "some," and "every" (or their equivalents in French, German, or any other language). But how can an assumption of this kind be defended? I shall argue that at a crucial point it has to rely on ordinary people's intuitions of deducibility.

Note, however, that the term "intuition" here is not being used in the sense of Spinoza (1914), Bergson (1903), or Husserl (1911). It does not describe a cognitive act that is somehow superior to sensory perception. Nor, on the other hand, does it refer merely to hunches that are subsequently checkable by sensory perception or by calculation. Nor does this kind of intuition entail introspection,[1] since it may just be implicit in a spoken judgment. Its closest analogue is an intuition of grammatical well-formedness. In short, an intuition that p is here just an immediate and untutored inclination, without evidence or inference, to judge that p.

To avoid any reliance on intuitions in that sense, it would be necessary to show that the assumption in question (about the nature of deducibility) is defensible within some well-recognised system of scientific procedure. Transcendental (that is, Kantian) arguments are obviously too controversial for the purpose. So either this procedure has to be empirically based and inductive, or it has to depend on some appropriate metamathematical theorem. But, as it turns out, neither of these strategies can wholly succeed: at crucial nodes an appeal to ordinary people's intuitions is indispensable.

The empirical-inductive strategy offers us an account of logic (as in, Stich 1975) in which it is viewed as an adjunct to science in general rather than, like geometry, an adjunct to physics in particular. Such an applied logic is understood as the combination of a formal system with appropriate interpretative rules; and it is to be tested, we are told, by assessing the explanatory and predictive power of the total theory that results from meshing it with the theories of the several sciences. In this way, it seems, what are accepted as logical truths turn out just to constitute one type of component in the total holistic system of what is accepted as scientific truth. They seem as much beholden to experiment and observation for their warranty as are any other scientific discoveries.

However, this kind of hard-line positivism comes up against some serious difficulties, which preclude it from supplying an intuition-free validation of deductive logic. First, certain regulative principles for theory construction, such as ideals of comprehensiveness, consistency, and simplicity, have in any case to be granted a priori status, so that in the defence of this status at least some principles of reasoning may have to be conceded an intuitive warranty. Second, much of the reasoning for which we need a logically articulate reconstruction does not take place in science at all but in law or administration, and is concerned not with what is in fact the case but with what ought to be. Third, the same logical principles have to be applied within each piece of scientific reasoning about the relative merits of two or more hypotheses, so that if ever any hypothesis has to be given up in the face of adverse experience it is always a factual, rather than a logical one. For example, we cannot claim, as does Reichenbach (1944), that quantum physics constitutes a restriction on the range of application of classical two-valued logic as well as of classical mechanics, because it is only in accordance with shared logical principles that it would be fair to elicit and compare the differing experimental conseqences of classical mechanics and quantum theory.[2] Hence, so far as we treat the totality of acceptable scientific hypotheses as constituting a single holistic system, we also need a single set of logical principles. Fourth, logically true statements are statements that are true in all logically possible worlds, and the evidence of happenings in the actual world must thus fall far short of establishing them.

Moreover, so far as the epistemology of a particular discipline is obligated to endorse the criteria of evaluation that are generally accepted in practice by reputable investigators in the field, it is certainly the appeal to intuitions that deserves endorsement for applied logic rather than the empirical-inductive strategy. Applied logicians make no attempt at all to test out their theories empirically within the context of a project for the holistic systematisation of all knowledge. On all the issues that are much discussed – issues about modality (for example, Quine 1960), subjunctive conditionals (Lewis 1973), indirect discourse (Carnap 1947), relative identity (Griffin 1977), proper names (Kripke 1972), adverbs (Davidson 1966), and so on – an implicit or explicit appeal to intuition provides some of the vital premises for the applied logician's argument.

Nor are the prospects for a metamathematical justification of applied logic any better than those for an empirical-inductive one. Any system in which rules of derivation are specified in formal terms is said to be "sound" if under some interpretation for the formalism

of the system it can be proved that from true premises these rules lead only to true conclusions. So it might seem as though, by thus using a semantic definition of logical consequence to check on a syntactic one, the rationality of a set of inferential rules could be established by experts in a metamathematical proof, without any recourse to intuitions other than those involved in the perception of the proof (Dummett 1978). But, though such a strategy has an agreeably professional appeal, it does not come to grips with the whole of the underlying epistemological problem. No reason is provided for supposing that the deductive liaisons of the logical particles of natural language can be mapped onto those of the connectives and quantifiers in the formal system that is proved to be sound.

For example, in any natural deduction system for the classical calculus of propositions the formula

1. $((A \rightarrow B) \,\&\, (C \rightarrow D))$

(or a notational variant) can constitute a premise from which

2. $((A \rightarrow D) \vee (C \rightarrow B))$

(or a notational variant) is derivable. And under the interpretation that Russell (1919) proposed for this calculus, a derivation could turn into an inference from

3. If John's automobile is a Mini, John is poor, and if John's automobile is a Rolls, John is rich

to

4. Either, if John's automobile is a Mini, John is rich, or, if John's automobile is a Rolls, John is poor

which would obviously be invalid. But what makes this invalidity obvious? The fact is that our own intuitions about the legitimate deductive liaisons of the logical particles (for example, the intuition that from the conditional "If John's automobile is a Mini, John is rich," we should be able to deduce the existence of a connection between antecedent and consequent that is independent of truth values) combine with our empirical knowledge of automobile costs to make it easy to imagine situations in which (3) is true and (4) is false. So though the propositional calculus is demonstrably sound, it resists Russell's interpretation as a logic of everyday reasoning in which conditional sentences may have a role, because it cannot capture intuitions like those on the basis of which we judge an inference from (3) to (4) to be invalid.[3] Admittedly, those intuitions might be said just to concern the *meanings* of the logical particles "if," "and," and "or," and there is nothing particularly remarkable, it might be objected, about the fact that one has to understand the meaning of an utterance to be able to appraise its validity. But the relevant point is that knowing the meanings of "if," "and," and "or" is indistinguishable from knowing, in principle, their legitimate deductive liaisons. So we cannot avoid appealing to intuitions of inferential validity in order to determine the claim of an interpreted formal system to constitute a theory of deducibility for everyday reasoning.

In other words, the problem of justification takes two rather different forms in regard to theories of deducibility. On the one hand, there is the issue of the theory's soundness, on the other, the issue of its application. Intuitions of inferential validity supply data in relation to the latter issue, not the former. But these intuitions are nevertheless an indispensable type of evidence for any theory of deducibility in everyday reasoning. Unless we assume appropriate intuitions to be correct, we cannot take the normative theory of everyday reasoning that they support to be correct. No doubt two different people, or the same people on two different occasions, may sometimes have apparently conflicting intuitions. But such an apparent conflict always demands resolution. The people involved might come to recognise some tacit misunderstanding about the terms of the problem, so that there is no real conflict; or they might repudiate a previously robust intuition, perhaps as a result of becoming aware that an otherwise preferred solution has unacceptable implications; or they might conclude that different idiolects or conceptions of deducibility are at issue.[4]

2. Intuitions as the basis of normative criteria for the evaluation of probability judgments. The position in regard to normative theories of probabilistic reasoning is rather analogous. We can take the mathematical calculus of chance, as axiomatised by Kolmogorov (1950), Reichenbach (1949), or Popper (1959a), to be a formal system that is open to semantical interpretation as a theory of the constraints that probability judgments of certain kinds *ought* to place on one another. But to just what kinds of probability judgment does the theory apply? This question has been much discussed. For example, proofs or arguments are available (Ramsey 1931; de Finetti 1931) to show that where probabilities are measured by betting quotients within a suitably coherent system of wagers their system conforms to the calculus of chance. A similar conformity has been demonstrated by Reichenbach (1949) and von Mises (1957) for the conception of probability as a relative frequency; by Carnap (1950) for the conception of probability as a type of logical relation that varies in strength along a spectrum that extends from contradiction at one extreme to entailment at the other; by Popper (1959b; 1968)and Mellor (1971) for the conception of probability as a causal propensity – a causally rooted tendency – and so on.

But none of those proofs or arguments establishes which conceptions of probability are operative – and under what conditions – in the everyday reasoning of lay adults, such as are the typical subjects of experiments carried out by cognitive psychologists. That is to say, it is one thing to establish one or more probabilistic interpretations for the calculus of chance, and quite another to show that the resultant theory applies to some or all of the probability judgments that are made in everyday reasoning. In order to discover what criteria of probability are appropriate for the evaluation of lay reasoning we have to investigate what judgments of probability are intuitively acceptable to lay adults and what rational constraints these judgments are supposed to place on one another. We have to select the conception or conceptions of probability in terms of which the most coherent account of lay judgments can be given, rather than evaluate those judgments by some single independently established standard.

The importance of this selection should not be underestimated. There are at least four ways in which it can make a lot of difference.

First, where probabilities are measured by betting quotients or construed as logical relations, we have to say – properly speaking – that they are functions of propositions; where they are relative frequencies, they are functions of sets; and where they are causal propensities, they are functions of properties (Cohen 1977b). Such categories of functions differ considerably in regard to their appropriateness for the evaluation of definite singular instances, as has often been pointed out (see Reichenbach 1949, pp. 376–77; Carnap 1950, pp. 226–28; Nagel 1939, pp. 60–75). They differ also in regard to their appropriateness for counterfactual inference (Cohen 1977b, pp. 306–9).

Second, where a probability is measured by a betting quotient, its statement is normally treated as an assertion about the strength of the speaker's belief in the outcome. Such a subjective fact is logically quite consistent with another speaker's having a different strength of belief in relation to the same issue. So when two people measure the probability of the same outcome subjectively by different betting quotients, they are not contradicting one another, whereas assertions of different relative frequencies, different logical relations, or different causal propensities would be logically inconsistent if they concerned the same issue.

Third, different probability functions may legitimately be assigned different values in relation to the same situation of uncertainty. Carnap, for example, demonstrated the existence of a nondenumerably infinite number of different measures for his logical-relation type of probability; the odds that are taken as appropriate to betting on a particular outcome on a given occasion need not correspond with the actual frequency of such outcomes in the relevant population; and the high frequency of B's among A's may be due to a series of coincidences, so that as a measure of causal propensity $p(B|A)$ [i.e. $p(B$ given $A)] = p(B)$ even though as a measure of relative frequency $p(B|A) > p(B)$. There is no mathematically demonstrable reason, therefore, why people should not in fact use different measures of probability for different situations or purposes, just as traders sometimes find it worthwhile to measure quantities of apples by weight and sometimes by number.

Fourth, if one or more semantic characterisations of probability are possible (as distinct from an implicit definition in terms of a set of mathematical axioms), then one might even need other formal systems than the calculus of chance to represent the syntax of some semantically defined categories of probability judgments other than the four already mentioned. The mathematics of probability may have no more reached its apogee in the work of Kolmogorov, than the mathematics of space did in that of Euclid. Nonclassical theories of probability may turn out to have an interest analogous to that of non-Euclidean geometries.

3. The systematisation of normative intuitions. It has been argued so far that any normative analysis of everyday reasoning – any statement that such and such lay judgments of deducibility or probability are correct, or incorrect, as the case may be – must in the end rely for its defence on the evidence of relevant intuitions. You cannot dodge this by an appeal to textbooks of logic or statistics. Of course, on any issue that can be settled empirically we naturally treat intuitions only as hunches that either will be confirmed by favourable observation or will give way to counter-observations. And in some area, such as the grammar of natural language, the question whether ultimate data are observational or intuitive or both is currently controversial: compare Chomsky (1965) and Sampson (1975). But on indisputably normative issues – on issues about how people *may* or *ought to* think or behave, as distinct from how they *do* – we cannot expect a major point at stake to be settled by observation. Here, if our aim is to build up a comprehensive system of theory, it is prudent to check our general hypotheses against intuitions in concrete individual cases – though in order to avoid an obvious risk of bias, these must always be the intuitions of those who are not theorists themselves. For example, the practice of the courts provides much evidence for a theory of lay intuitions about probability in forensic reasoning (Cohen 1977b), but writers on this subject should not invoke their own intuitions.

Normative theories are subject to the usual inductive criteria. They are better supported if they apply to a wider rather than a narrower range of significantly different kinds of intuitive inference or judgment, just as the more comprehensively explanatory theories have greater merit in natural science. But there would obviously be a point at which even the mere process of putting problems to a person in varied contexts, in order to extract his intuitions, could reasonably be taken to cross over into a procedure for changing his normative outlook instead of just recording it. Thus, recent writers on ethics (for example, Rawls 1972; Daniels 1979; 1980) have distinguished between the narrow reflective equilibrium that is constituted by a coherent reconstruction of a person's existing moral principles, where only an occasional intuition is repudiated (for the sake of consistency), and the wide reflective equilibrium that is obtained when a person chooses between his existing moral principles and proposed alternatives, on the basis of sociological, historical, economic, psychological, or other considerations that may weigh with him. In matters of deducibility or probability the analogue of this philosophical choice would occur in the process of education, research, or philosophising whereby hitherto uncommitted students are sometimes transformed into thoroughgoing Quineians, say, or Bayesians, or Popperians, so that they come to adopt substantially different conceptions of deducibility or probability from those once operating in their untutored judgments. But the normative theories that are at issue in the present context require a narrow, not a wide, reflective equilibrium. The judgments of everyday reasoning must be evaluated in their own terms and by their own standards.

Nevertheless, even at the level of narrow equilibrium, generality is only purchased at a cost. In constructing any normative theory the same strategy of abstraction and idealisation that we find in natural science is appropriate as a trade-off for the purchase of increased

comprehensiveness. A figure was sought, for instance, in classical mechanics for the acceleration of a falling body. But it would have been too difficult to construct a satisfactory general hypothesis about this figure unless it had been supposed to apply to falling bodies in abstraction from any frictional effect of the medium through which they are falling. So most general theories in natural science have come to be constructed in terms of idealised entities and precisely measurable properties – particles in frictionless media, perfect gases, closed and isolated systems, lines without breadth, and the like – irrespective of the fact that these categories may not actually be instantiated anywhere in nature. Such a theory may nevertheless be used for the explanation or prediction of actual events if appropriate allowances are made for the extent to which, and the reasons why, they differ from the pertinent idealisation. Analogously, normative hypotheses achieve generality by a similar process of abstraction, idealisation and precisification that is highly important to bear in mind.

For example, the moralist says that promises should be kept. But implicitly or explicitly the moralist has to safeguard that generalreation against the awkward counterexamples – helping a potential murderer, and the like – where a promise is rightly broken. The unqualified norm must be regarded as holding good only in an ideally simplified world in which none of those awkward circumstances is ever present. In the same way, logicians construct theories of natural deductions in terms of propositions, which have truth values and entailments but no time or place or causal connexions. Consequently, since the actual judgments and reasonings of human beings occur on particular dates and in particular locations, in a particular causal context, a sequence of logical formulas can only be taken to represent an actual piece of human reasoning if due regard is paid to the various respects in which the representation inevitably constitutes something of an abstraction from – or idealisation of – its original.

4. The derivation of an account of human competence in deductive or probabilistic reasoning. If a physicist observes the position of the needle on a certain dial under chosen experimental conditions, then the datum to be explained is the position of the needle, not the fact that someone observes it. The event observed, not the act of observing it, is what is relevant. Otherwise optics (and perhaps acoustics) would be all the science that there is. Analogously, the datum that the moralist has to take into account is the rightness or wrongness of a particular action, not the deliverance of conscience that pronounces it right or wrong; and the logician's datum is the validity or invalidity of a particular inference, not the intuition that assures us of it. So enquiry into the norms of everyday reasoning no more aims at a theory *about* intuitions than physics or chemistry aims at a theory *about* observations. Epistemology does not dominate ontology here. And fortunately it is not necessary for present purposes to determine what exactly the study of moral value, probability, or deducibility has as its proper subject matter. For example, an applied logician's proper aim may be to limn the formal consequences of linguisitic definitions (Ayer 1946), the most general features of reality (Quine 1960), or the structure of ideally rational belief systems (Ellis 1979). But, whatever the ontological concern of applied logicians, they have to draw their evidential data from intuitions in concrete, individual cases; and the same is true for investigations into the norms of everyday probabilistic reasoning.

It follows that for every such normative theory, which determines how it is proper to act or reason, there is room to construct a factual theory that does take intuitions as its subject matter. This factual theory will describe or predict the intuitive judgments that formulate the data for the corresponding normative theory. It will be a psychological theory, not a logical or ethical one. It will describe a competence that normal human beings have – an ability, uniformly operative under ideal conditions and often under others, to form intuitive judgments about particular instances of right or wrong, deducibility or nondeducibility, probability or improbability. This factual theory of competence will be just as idealised as the normative theory from which it derives. And though it is a contribution to the psychology of cognition it is a by-product of the logical or philosophical analysis of norms rather than something that experimentally oriented psychologists need to devote effort to constructing. It is not only all the theory of competence that is needed in its area. It is also all that is possible, since a different competence, if it actually existed, would just generate evidence that called for a revision of the corresponding normative theory.

In other words, where you accept that a normative theory has to be based ultimately on the data of human intuition, you are committed to the acceptance of human rationality as a matter of fact in that area, in the sense that it must be correct to ascribe to normal humans beings a cognitive competence – however often faulted in performance – that corresponds point by point with the normative theory. Of course, it would be different if you believed in some other source of normative authority. If, for instance, you believe in a divinely revealed ethics, you are entitled to think that some people's competence for moral judgment may fall short of correct moral ideals: you could consistently invoke some doctrine of original sin to account for the systematic failure of untaught intuition to accord with the correct norms of moral judgment. But, if you claim no special revelation in ethics, you will have to take intuitive judgments as your basis, and then people's competence for moral judgment – as distinct, of course, from their actual performance in this – cannot be faulted. Analogously, if you claim no special revelation in matters of logic or probability, you will have to be content there too to accept the inherent rationality of your fellow adults.

To ascribe a cognitive competence, in this sense, within a given community is to characterise the content of a culturally or genetically inherited ability which, under ideal conditions, every member of the community would exercise in appropriate circumstances. It states what people can do, rather than what they will do, much as the characterisation of a linguistic competence can be taken to describe what it is that native speakers must be assumed capable of recognis-

ing about the structure of morphophonemic strings (Chomsky 1965) rather than what they do actually recognise. The fact is that conditions are rarely, if ever, ideal for the exercise of such a competence. Just as passion or self-interest may warp our moral discernment, or memory limitations may restrict the length of the sentences we utter, so too a variety of factors may interfere with the excercise of a competence for deductive or probabilistic reasoning. A local unsuitability of childhood environment may inhibit the maturation of innate ability, education (that is, education in subjects other than logic and probability theory) may fail to make the most of it, individual disabilities or normal memory limitations may set limits to what even the best environment and education can achieve, and various motivational and other factors may operate to induce malfunctions of the relevant information-processing mechanisms. To suppose that all normal adults are able to reason deductively is certainly not to suppose that they will never err in their judgments of logical validity, and still less that they will in practice execute any particular finite chain of reasoning that is called for, however complex it may be, just so long as it is licensed by intuitively evident rules of natural deduction. In practice, our rationality is "bounded" (Simon 1957, pp. 198–202). We are all able to walk, if in normal health, but it does not follow that we can all walk on a tight rope, or for a thousand miles without stopping, or that none of us ever stumbles. It is here that the issues arise that will be discussed in the second part of this paper.

In short, accounts of human competence can be read off from the appropriate normative theories, so far as they are based on the evidence of intuitions; accounts of actual performance under different conditions are to be obtained by experiment and observation; and hypotheses about the structure and operation of human information-processing mechanisms must then be tested against the facts of competence and performance that it is their task to explain. The structure or design of such a mechanism must account for the relevant competence, but its operation must be subject to various causes of malfunction that will account for the flaws found in actual performance.

One may be tempted to ask: "How do we know that any intuition of the relevant kind is veridical?" But to ask for knowledge here is to ask for what is in principle impossible, at least in the sense in which knowledge is something like justified true belief, and where there is no alternative to invoking intuition, since an intuitive judgment that *p* essentially lacks any external ground to justify accepting that *p*. The best that normative theorists can hope for in this field (and also what they need to achieve), if they do not claim any special revelation, is that the contents of all relevant intuitions – suitably sifted or qualified, if necessary – can be made to corroborate one another by being exhibited as the consequences of a consistent and relatively simple set of rules or axioms that also sanctions other intuitively acceptable, but previously unremarked, patterns of reasoning. The inductive principle of mutual corroboration here is analogous to that operative in natural science, as Bacon long ago pointed out in regard to normative theories of ethics or jurisprudence (Kocher 1957; cf. Cohen 1970).

It would be different if we were evaluating the cognitive competence of some other species, or even of human children. We should be free to find their intuitive efforts at probabilistic reasoning, for example, or their moral sensitivity, to be rather inferior by the standard of our own norms. But we cannot attribute inferior rationality to those who are themselves among the canonical arbiters of rationality. Nothing can count as an error of reasoning among our fellow adults unless even the author of the error would, under ideal conditions, agree that it is an error.

Other arguments about rationality do not concern us here. It is true that, even where animals, children, or Martians are concerned, there are limits to the extent to which we can impute irrationality. As has been well remarked (Dennett 1979, p. 11), if the ascription of a belief or desire to a mouse is to have any predictive power, the mouse must be supposed to follow the rules of logic insofar as it acts in accordance with its beliefs and desires. But this is not to suppose that the mouse has any great powers of ratiocination. Equally (Quine 1960) we have to impute a familiar logicality to others if we are to suppose that we understand what they say: different logics for my idiolect and yours are not coherently supposable. But there is always the possibility that we understand less than we think we do and that some imputations of logicality are therefore not defensible on this score. Again, evolutionary pressures in the long run eliminate any species that is not sufficiently well equipped to surmount threats to its biological needs. But evolutionary considerations are better fitted to put an explanatory gloss on the extinction of a species after this event has already occurred than to predict the precise level of rationality that is required for this or that species' continued survival within its present environment.

What I have been arguing is that normative criteria for ordinary human reasoning rely for their substantiation on a procedure analogous to what is called "bootstrapping" in artificial intelligence (see Dawes & Corrigan 1974). The intuitions of ordinary people are the basis for constructing a coherent system of rules and principles by which those same people can, if they so choose, reason much more extensively and accurately than they would otherwise do.[5] Consequently these ordinary people cannot be regarded as intrinsically irrational in regard to any such cognitive activity. An investigator who wanted to make out a serious case for deep-level human irrationality in this area might be tempted to operate with normative criteria that were the product of philosophical argument for some appropriately wide reflective equilibrium and consequently differed from the narrow, bootstrapping, reflective equilibrium which merely reconciles intuitions. But any kind of scientific or mathematical reasoning to which such criteria directly apply has a specialised and technically regimented quality that makes it difficult or impracticable for those who have not been trained appropriately. For example, it may involve deduction within an artificial language system, or employ relatively sophisticated concepts of statistical theory.

Hence the investigator's experiments would founder in a characteristic indeterminacy. They would constitute an accurate test of their subjects' competence for reasoning only to the extent that these subjects were not ordinary people but specially trained experts. So the results of the test might reveal how good was the training or how effective were the procedures for selecting people to be trained; they would tell us nothing about the rationality or irrationality of untrained people. Though a person may well acquire a wide reflective equilibrium with regard to ethical issues that is inconsistent with a previously existing narrow reflective equilibrium, there is no possibility of an analogous inconsistency with regard to deducibility or probability. In the case of deducibility, narrow reflective equilibrium remains the ultimate framework of argument about the merits of other deductive systems, and in the case of probability, we are merely replacing some modes of measuring uncertainty by others.

II. Four categories of research into defects of cognitive rationality

The past decade or so has seen the growth of a vast literature of psychological research into replicable defects of human reasoning. Often investigators are content just to argue for the existence of such defects and to suggest explanations. But sometimes they also claim justification for extensive criticisms of human rationality. Several convenient reviews of this literature are already available (for example, Slovic, Fischhoff & Lichtenstein 1977; Nisbett & Ross 1980; Einhorn & Hogarth 1981), and I do not aim to produce another here. Rather, the purpose of the second part of the paper is to establish four categories, into one or the other of which, on close assessment, any item in this literature may be seen capable of being assigned without "bleak implications for human rationality," once an account of the normative criteria for ordinary human reasoning is agreed to entail an ascription of the corresponding competence to ordinary human adults (as argued in part I). For reasons that will emerge in the sequel, these four categories of research activity are appropriately entitled "Studies of cognitive illusions," "Tests of intelligence or education," "Misapplications of appropriate normative theory," and "Applications of inappropriate normative theory." Examples will be furnished for each category. It will be assumed in every case that the phenomena reported are replicable, and that no technical faults occur in the presentation of the data, such as miscalculations of statistical significance: if the examples furnished here are in fact faulty, others are easily found. The issues raised here do not concern the robustness of the phenomena, solely their interpretation.

The categorisation is intended to be an exhaustive one, not in the sense that every item in the literature is actually assigned to one of the four categories but that in principle it could be. Claims that human reasoning tends to be invalid in certain circumstances are either correct or incorrect. The correct claims relate either to fallacies that, on reflection, everyone would admit to be such, as in studies of cognitive illusions, or to fallacies that require some more elaborate mode of demonstration as in tests of intelligence or education; the incorrect claims result either from misapplications of appropriate normative theory or from applications of inappropriate normative theory.

1. Studies of cognitive illusions. In view of what has been argued above about ordinary people's competence for deductive and probabilistic reasoning, there is a prima facie presumption, in regard to any experimental data in this area, that they can be explained as a manifestation of some such competence, even though the details of the explanation may not be easy to fill in. Where no explanation of this kind is available, one possibility is that experimenters have created a cognitive illusion. They have manipulated the circumstances of a situation in such a way that subjects are induced to indulge in a form of reasoning that on a few moments' prompted reflection they would be willing to admit is invalid.

A very good example of this is the familiar four-card problem (Wason 1966). The subjects are presented with four laboratory cards showing, respectively, 'A,' 'D,' '4,' and '7,' and know from previous experience that every card, of which these are a subset, has a letter on one side and a number on the other. They are then given this rule about the four cards in front of them: "If a card has a vowel on one side, then it has an even number on the other side." Next they are told: "Your task is to say which of the cards you need to turn over in order to find out whether the rule is true or false." The most frequent answers are "A and 4" and "only A," which are both wrong, while the right answer "A and 7" is given spontaneously by very few subjects.

Wason and his colleagues, in attempting to account for these data (see Johnson-Laird & Wason 1970), interpret the error as a bias towards seeking verification rather than falsification in testing the rule. But if that were the nature of the error one would expect "D" to show up in the answer like "4" does, since the contrapositive equivalent of the rule is, "If a card does not have an even number on one side, it does not have a vowel on the other." Perhaps it will be said that this is simply owing to a failure to grasp the equivalence of contrapositives here. But such a failure would account also for the absence of "7" from most answers, without the need to suppose that in testing a conditional rule subjects do anything other than check, in each case in which the antecedent holds true, whether the consequent does also: this is because the presence of "4" in many answers may then be put down to the prevalence of inference from an utterance that is of the form "if p then q" to an utterance that is of the form "if q then p" – a prevalence for which there is independent evidence (see II.3 in regard to the fallacy of illicit conversion). I shall assume, therefore, that the subjects' specific error here is best interpreted as a failure to apply the law of contraposition. What then causes that failure?

It would be wrong to conclude that the deductive competence of most logically untutored subjects does

not embrace the law of contraposition. A subsequent experiment (Wason & Shapiro 1971) has been claimed to show that if the four cards are those related to a more concrete rule, namely, "Every time I go to Manchester, I go by train," then substantially more subjects are successful. Even better results were obtained (Johnson-Laird, Legrenzi & Sonino Legrenzi 1972) when the rule was, "If a letter is sealed, then it has a fivepenny stamp on it" and the laboratory cards were replaced by (sealed or unsealed, stamped or unstamped) envelopes. Further experimentation (Van Duyne 1974; 1976) has also been claimed to show that degrees of realism in fact affect performance in a continuous, linear way. However, it looks as though we need to distinguish here between two different ways in which realism may be increased. One is by writing descriptive words or sentences on the cards, instead of just letters and numerals, and altering the rule accordingly: the other is by using real objects (envelopes) instead of cards. The results of Manktelow and Evans (1979) suggest that when realism is increased in the former manner the fallacy still occurs. But those results do not weaken the finding that when real objects replace cards the fallacy hardly ever occurs. It seems, therefore, that experimenters' power to generate an illusion here depends on the relative unfamiliarity and artificiality of their apparatus. In their familiar concrete concerns human beings show themselves well able to apply the law of contraposition to appropriate problems. Faced instead with a situation in which the items against which a conditional rule is to be checked are things (cards bearing letters, numerals, words, sentences, geometrical diagrams, and the like) that echo the symbolism in which the conditional rule itself is formulated, subjects' reasoning tends to be led astray in the "matching bias" to which Manktelow and Evans have traced the fallacy.

The point of describing experimental effects like Wason's as cognitive illusions is to invoke the analogy with visual illusions: it is in no way intended to derogate from their importance, nor to suggest that if the circumstances that cause the illusion occur naturally (as distinct from being the result of an experimenter's contrivance) then the illusion will not occur.[6] The discovery of any such effect in human performance generates a significant piece of evidence about the way in which the underlying information-processing mechanism operates. The findings about the four-card problem may legitimately be said (Johnson-Laird & Wason 1977) to support the view (Piaget 1972) that most people manage to apply their logical competence without ever formulating it expressly at a level of generality sufficient for it to be readily applicable to wholly unfamiliar tasks. People will distinguish form from content in their reasoning, or extrapolate accurately from one content to another, only to the extent that similarity of form is accompanied by some rough equality of vital interest. So subjects who reason fallaciously about the four-card problem need not be supposed to lack the correct deductive "program." Indeed, none of the experimenters in the area suggests this. The subjects merely fail to recognize the similarity of their task to those familiar issues in which they have profited by using the deductive procedure of contraposition. As a result, either that procedure receives no input or its output is deleted, and the behaviour of the subjects manifests a matching bias.

Analogously, other experimental data show, it has been argued (Wason 1960; 1968), that in an abstract task, like hypothesising about the rule that generates a given series of numbers, most people are unable to use the procedure of proving a hypothesis by eliminating alternatives to it.[7] They tend to seek confirmatory evidence for their favoured hypothesis rather than disconfirmatory evidence for alternatives to it. In addition, they often do not relinquish hypotheses that have been shown to be false. Nor is people's eliminative performance substantially better when confronted with a computer-screen simulation of a simple mechanical problem (Mynatt, Doherty & Tweney 1977). Yet it hardly needs an experiment to show that most people are quite capable of using eliminative procedures correctly when dealing with real objects – not simulated ones – in familiar everyday situations: if the soap is not in the basin, we reason, it must be in the bath; if one's caller has not come by automobile, he must have walked; and so on. So it is not that ordinary people lack competence for the kind of deductive inference that moves from "*p* or *q*" and "not-*p*" to "*q*," which is essential to all such eliminative reasoning. It seems rather that in normal investigative situations the disjunctive premise for this pattern of reasoning is supplied by previous experience, and in an artificial or unfamiliar situation we lack the relevant kind of previous experience to supply the input. To build up that experience some pursuit of confirmatory (as distinct from disconfirmatory or eliminative) strategies would not be unreasonable, as is recognised by Mynatt, Doherty, and Tweney (1978, p. 405); and retention of a falsified hypothesis would even be desirable if it explained quite a lot of the evidence and no unfalsified hypothesis were available that had as good explanatory value. So too the sharply falsificationist model of scientific progress that was originally offered by Popper (1959a) has rightly met with substantial criticism from other historians and philosophers of science (for example, Swinburne 1964; Lakatos 1970).

Again, it may seem puzzling that subjects seem unable to judge correctly in the laboratory that some event is controlled by another, or is independent of it, as the case may be, and yet the very same subjects get along all right most of the time in their everyday life (Jenkins & Ward 1965). But the puzzle may be lessened by considering some of the ways in which their experimental tasks are not representative of the normal conditions for such judgments: check-ups are excluded, temporal variations are absent, the output considered is unnaturally discrete, the response has to be a relatively hurried one, and so on.

Experimenters need to devise a great variety of experiments involving such cognitive illusions in order to test out theories about how human beings in fact perform, or fail to perform, the various acts of reasoning for which they apparently possess a competence. But one should recognise these experiments for what they are, and not conceive of the illusions that they generate as some kind of positive, though fallacious, heuristic that is employed by the subjects. Consider, for

example, the supposed heuristic of availability (Tversky & Kahneman 1973). A person is said to employ this heuristic whenever he estimates frequency or probability by the ease with which instances or associations are brought to mind. For example, if student subjects, when asked in the laboratory, erroneously judge English words beginning with *re* to be more frequent than words ending with *re*, they are diagnosed to have employed the heuristic of availability because the former words are more easily brought to mind than the latter. But a heuristic is a way of finding something out that one does not already have at the front of one's mind. The availability illusion consists instead in relying on data that one already has at the front of one's mind. There is a lot of evidence that most people are too slow to change certain kinds of beliefs (Ross & Lepper, forthcoming). But no one thinks that this evidence establishes a "conservatism heuristic," rather than that it just manifests the influence of factors which make for belief inertia. Analogously, if the argument for rational competence (in part I) is accepted, the "availability" results must be interpreted to have shown, not that the subjects are estimating the frequency or probability of an x by reference to the availability of an x, but that they are doing this by reference to those x's that happen to be available. The subjects are not to be construed as operating on the evidently wild assumption that frequency can safely be taken to equal availability. Rather, where A is the available population, they are to be construed as operating on the not so evidently wild assumption that frequency can safely be taken to equal frequency in A, which is a very different matter.

In other words, to be entitled to recognise an error in subjects' reasoning here, we have to attribute to them a conception of the frequency or probability of an x, $p(x)$, such that it is incorrect to infer $p(x) = n$, where x's are y's, from $p(x \mid y)$ [$p(x$ given $y)$] $= n$, unless the y's are a suitably representative sample of the total population. So we thereby (see part I) also attribute to them the competence to avoid those incorrect inferences. It follows that their probability-estimating mechanism must be supposed to include some such procedure as: check whether available evidence constitutes a fair sample in relevant respects, and, if not, seek evidence that is of the missing kind or kinds. What happens is just that the operation of this procedure tends to be obstructed by factors like the recency or emotional salience of the existing evidential input, by the existence of competing claims for computing time, or by a preference for least effort. Cognitive illusions, in the laboratory or in real life, depend on the power of such factors to hold subjects back, under the pressure of interrogation, from obtaining an appropriate additional input to their information-processing operation, just as when a visual conjurer relies at a crucial moment on his own speed of action, and on the visual inattentiveness of those who are watching, to hold the latter back from obtaining an appropriate additional input to their visual information-processing operation.

Another procedure referred to in the literature as a "heuristic" is the method of anchoring and adjustment, whereby a natural starting point or anchor is used as a first approximation to the required judgment of frequency, probability, expected value, and so on, and is then adjusted to accommodate the implications of additional information. Tversky and Kahneman (1974) have shown the existence of a tendency for adjustments to be insufficient: subjects with high starting points end up with higher estimates than those with lower ones. This tendency has also been noted (Lichtenstein & Slovic 1973) in an experiment with people who were gambling in a Las Vegas casino. Even there, an element of conjuring was present, in that the game that was played was specially designed for the purpose. But it would obviously be implausible to suppose that any kind of cognitive illusion occurs only when the circumstances that cause it are deliberately contrived: conjuring is not the only source of visual illusions either.

A somewhat similar phenomenon has been demonstrated in relation to hindsight (Fischhoff 1975): judges with knowledge of the outcome tend to overestimate the probability that they would have declared prior to the event. This is like starting with an anchor at 100% probability and adjusting to allow, not for more information, but for less, that is, for ignorance of the actual outcome.

However, unlike in the case of the supposed heuristic of availability, there is nothing intrinsically fallacious in the procedure of anchoring and adjustment. It is a perfectly legitimate heuristic if correctly operated. What goes wrong is just that the effects of recency or salience are generally too strong to permit correct operation. Thus Slovic et al. (1976) are right to point out that bias from anchoring, like that from availability, is congruent with the hypothesis that human reasoners resemble computers that have the right programs and just cannot execute them properly. But Slovic et al. also claim that there are certain other errors prevalent in probabilistic reasoning, concerned with sampling and prior probabilities that are not congruent with this hypothesis, and we shall see shortly that the latter claim cannot be sustained.

2. Tests of intelligence or education. A second category of research activity found in the literature concerns ignorance, not illusion. It demonstrates a lack of mathematical or scientific expertise.

A lack of mathematical expertise here amounts to an ignorance of principles that not everyone can be expected to acknowledge readily, still less to elicit spontaneously from their relevant competence. Possession of a competence for deductive or probabilistic reasoning entails the possession of a mechanism that must include not only certain basic procedures, corresponding to a set of axioms or primitive rules for the normative system concerned, but also a method of generating additional procedures, corresponding to the proof of theorems or derived rules in that normative system. But the actual operation of this method, beyond its simplest forms, may require skills that are relatively rare, just as a particular talent is required for the discovery of proofs in logic or mathematics wherever no mechanical decision procedure is known. In the latter case what are needed in an outstanding degree are such capacities as those for discerning shared structure in superficially different materials, for memorising complex relationships, and the like – in

other words, whatever promotes the proposal of worthwhile hypotheses in the trial-and-error search for appropriate connexions. Correspondingly, only people with those skills in an outstanding degree can be expected to generate interesting new procedures for eliciting deductive consequences or estimating probabilities. Only they will be able to supply spontaneously the input, in terms of perceived similarities and the like that will enable the method of generating additional procedures to operate fruitfully. Others will have to learn these proofs or derivations, or acquire the additional procedures, at second hand. Education must supplement innate intelligence, where intelligence is understood not as the competence that everyone has but as the level of those skills that are required to supply the novel input essential for the discovery of proofs. So experiment in this area may be able to show us the limits of ordinary people's intelligence, in the appropriate sense, or the extent to which subjects have profited from logical or mathematical education. But it cannot demonstrate an erroneous competence.

For example, it required the genius of a great mathematician (Bernoulli 1713) to discover and prove that, if you estimate the probability of a certain characteristic's incidence in a population from its frequency in a sample, then the probability of your estimate's being correct, within a specifiable interval of approximation, will vary with the size of the sample. So it is easily understandable that psychological experiment finds a tendency among ordinary people, untutored in statistical theory, to be ignorant of this principle and its applications (Tversky & Kahneman 1971). No doubt equally cogent experiments could be designed to establish the fact that those untutored in Euclidean geometry are still ignorant of the fact that the square on the longest side of a right-angled triangle is equal in area to the sum of the squares on the other two sides, since it required another outstanding mathematician, Pythagoras, to discover a proof of this fact. Again, it is said (Tversky & Kahneman 1971, p. 109) that at a meeting of mathematical psychologists and at a general session of the American Psychological Association the typical respondent attached excessive significance to inferences from relatively small samples. But what this adds to the previous finding is just a reason for reassessing the extent or success of the education that the respondents had in fact undergone. And the same holds true in relation to those who are supposed to have some statistical training but still fail to recognise new examples of regression to the mean for what they are (Kahneman & Tversky 1973).

Not all errors of estimation that are due to ignorance arise from subjects' deficiencies in mathematical expertise. Some arise instead from subjects' deficiencies in scientific (for example, psychological) expertise. For example, there is a good deal of evidence (reviewed in Slovic et al. 1977, pp. 5–6) that people are often overconfident in their second-order estimates of the accuracy of their own primary estimates. What happens here is that they are unaware of the various ways in which the information-processing mechanism that generates the primary estimates may be affected by performance error. However, this is scarcely surprising, since the facts about those patterns of error are being discovered only gradually and only by difficult (and sometimes controversial) research. No doubt it would be salutary if all nonpsychologists were taught every such fact that has been properly established. But all that is discovered, when their ignorance of such a fact is discovered, is a gap in their education.

3. Misapplications of appropriate normative theory. We have been concerned so far with genuine fallacies, to which experiments reveal that people are prone, because of either illusion or ignorance. However the literature also contains several examples of more questionable claims that a common fallacy exists. These are situations in which the experimental data may be explained as a direct manifestation of the relevant competence without any need to suppose an error in performance. Such claims arise either through a misapplication of the appropriate normative theory or through an application of an inappropriate one.

One particularly instructive example of the former kind relates to the alleged prevalence of the fallacy of illicit conversion, and, in particular, of inference from a proposition of the form "if p then q" to one of the form "if q then p." Intellectuals have remarked for over two millennia (Hamblin 1970) on the tendency of their inferiors to commit this fallacy, and in recent years it too has been a topic for psychological investigation (Wason & Johnson-Laird 1972). The investigators conclude that, in situations in which subjects are apparently prone to illicit conversion of conditionals, "this is not because the subjects possess faulty rules of inference but because they sometimes make unwarranted interpretations of conditional statements" (p. 65). The subjects are claimed to treat these conditionals as if they were statements of causal connexion which allow one to infer from effect to cause as well as from cause to effect.

But it is not clear that the subjects must in fact be supposed even to be making an unwarranted interpretation. We have to bear in mind here that the principles of a normative theory, such as one that systematises criteria for deducibility, inevitably involve abstraction and idealisation (see part I, section 3). So what are to be taken as the actual, concrete premises that are represented by the initial formulas in a primitive or derived rule for natural deduction, when such rules are taken to be the norms relevant to some actual sequence of human reasoning? The mere sentences uttered do not normally constitute all of the premises conveyed by the total act of communication, since we are presumptively entitled to take the latter as including also any judgments that are implied by the act of uttering those sentences in the contextual circumstances. For example, as far as human conversation is governed by rules of relevance, brevity, informativeness, and so on, as required by the purpose in hand (Grice 1975),[8] the information provided by the utterance of a solitary conditional sentence – if p then q – may be presumed, unless there are specific indications to the contrary, to be all that is required in the circumstances to satisfy the interest either of someone who wants to know what is also true if the antecedent

of the conditional is true, or of someone who wants to know the conditions under which the consequent of the conditional sentence is true. In the former case ("If you interrupt him now, he'll be cross") the conditional is convertible because its utterance would normally be pointless unless "if not-*p* then not-*q*" were also true and "if not-*p* then not-*q*" is formally equivalent to the converse of "if *p*, then *q*." In the other case ("If you give him a tip, he'll let you in") the conditional is convertible because its solitary utterance may be presumed to state the only condition under which the consequent is true.

Hence if we consider the total content of the message communicated, rather than just the conditional sentence that is uttered, it would not be fallacious or unwarranted for subjects to presume, unless there are specific indications to the contrary, that the converse of the conditional is implicit in the message, and the convertibility of causal conditionals is just a special form of this. A psychological experimenter who wishes to exclude the legitimacy of presuming the converse in such a case must contrive suitable instructions to his subjects and teach them how to distinguish between the implications of a sentence uttered and the implications of its utterance. But how could we judge the suitability of such instructions without taking into account the extent of their success in averting inferences to the converse? In other words, a tendency to commit the fallacy of illicit conversion in everyday life is demonstrable only on the basis of an unrealistic assumption – namely, that when a normative theory is invoked for the evaluation of commonsense reasoning its criteria should be applied to nothing but the linguistic forms that are actually uttered.

Another line of research activity (see, for example, Wagenaar 1972) in which appropriate norms seem to be sometimes misapplied is in studies of judgments of randomness. Results over quite a variety of tests seem to confirm the hypothesis that subjects who are attempting to behave randomly will produce series that have too many alternations and too few repetitions. But as has been well pointed out (Lopes 1980), a series may have randomness with respect to its atomic or elementary events, while still possessing molecular units, such as groups of ten consecutive atomic events, that do not exhibit randomness. Or randomness may be achieved for a certain category of molecular events, at the cost of sacrificing randomness with respect to elementary events. Unless this distinction between different kinds of randomness is clearly presented to the subjects, they are not in a position to know what kind is being sought by the experimenters. And again it is not easy to see how the subjects' apparent failure to produce correct judgments of randomness should not be regarded as simply a measure of the aptness of their instructions.

A different way in which an appropriate normative theory may be misapplied was instantiated in the course of an attempt to show that, as compared with their treatment of predictive evidence, people are prone to "a major underestimation of the impact" of diagnostic evidence "which could have severe consequences in the intuitive assessment of legal, medical, or scientific evidence" (Tversky & Kahneman 1977, p. 186). Subjects were given two sets of questions that were regarded by the experimenters as similar in relevant structure. But in fact the predictive set concerned conditional probabilities, as in – for one instance – "The chance of death from heart failure is 45% among males with congenital high blood pressure," while the diagnostic set concerned unconditional ones, as in "The radiologist who examined Bill's X-ray estimated the chance of a malignancy to be 45%," and this difference sufficed to account quite rationally for differences in the numerical answers to the two sets of questions (see Cohen 1979, pp. 403–5). Moreover, when the dissimilarity of structure was remedied, the alleged phenomenon of diagnostic underestimation failed to emerge. Other results that appear to evince this phenomenon have to be discounted for different reasons (Cohen 1979, pp. 401–3). But the failure to distinguish appropriately here between conditional and unconditional probabilities is a good example of how the appearance of a fallacy in subjects' reasoning may be generated by a slip in the application of the appropriate probabilistic analysis,[9] since within normative probability theory the distinction between conditional and unconditional probabilities is well established.

Even the so-called gambler's fallacy, or "fallacy of the maturity of chances," which is sometimes referred to in the literature on cognitive irrationality (Tversky & Kahneman 1974; Hogarth 1975), comes under some suspicion. More empirical work seems necessary here, but there are at least three possible approaches to the phenomenon that call into question its interpretation as a fallacy of probabilistic reasoning.

If some people believe that after a long run of heads the probability of tails on the next toss will be greater than ½, then one possibility is that they should be interpreted as believing thereby in a spirit of distributive justice that regulates the whole cosmos with a policy that ensures ever-increasing probabilities of a trend-reversing intervention whenever identical outcomes begin to succeed one another within an otherwise chance set-up. On this construal, a gambler's metaphysical belief may be at fault, but not the rationality of his reasoning from it. However, such an interpretation needs independent evidence to support the attribution of belief in the particular case. Otherwise it is open to the charge of being culpably ad hoc, if not of merely repeating what is to be explained within the explanation.

Second, we may need to distinguish here between two rather different probabilities, either of which might be a matter for estimation. Is the gambler supposed to be estimating, in relation to the next toss of a fair coin, the probability of a tails outcome within a space that consists of the two alternative outcomes: heads and tails? Or is the probability in mind, at the *n*th toss of a fair coin, that of having at least one tails outcome within any space that consists of *n* outcomes? Whereas the correct figure for the former would be ½, the correct figure for the latter would get greater and greater than ½ as *n* itself increases beyond 1, in accordance with Bernoulli's theorem. To ascertain clearly and unmistakably which of the two probabilities is being estimated it would be necessary to question

the gambler in a way that would tend to discourage any incorrect estimate, since in order to convey the exact meaning of a particular type of probability assignment (or, indeed, of any other type of statement), one needs to state the conditions under which such a judgment is true. So we are left with a characteristic indeterminacy here. Any attempt to extract an exact answer from the gambler would transform the situation in a way that would tend to disconfirm the occurrence of fallacious reasoning, and to the extent that the situation was not so transformed, the exact nature of the situation would remain in doubt. But it remains an open question, in view of what was said earlier about ordinary people's ignorance of Bernoulli's theorem, whether ordinary gamblers may legitimately be expected to be aware of its implications.

Finally, it may be that the matter at issue needs to be regarded more as a pragmatic than as a cognitive phenomenon. In the long run a gambler could integrate the so-called fallacy into a winning strategy against any opponent who always insists on even odds but is willing to play as long as the gambler wants: the gambler has only to continue increasing the stakes sufficiently at each toss until tails actually comes up. But, of course, such a strategy could be executed only within the limits of any restriction that is imposed on the stakes either by the opponent or by the gambler's resources, just as any intellectual competence is subject to limitations in actual performance.

4. Applications of inappropriate normative theory. There is a tendency for some investigators of irrationality to proceed as if all questions about appropriate norms have already been settled and the questions that remain open concern only the extent of actual conformity to these norms. It is as if existing textbooks of logic or statistics had some kind of canonical authority. But in fact many important normative issues are still controversial. For example, it seemed at one time that at least the Frege-Russell logic of quantification had become a universally received doctrine. But its closeness of fit for the appraisal of natural-language reasoning is now under a powerful challenge (Sommers 1981) from work that exploits hitherto undiscovered ways of developing the Aristotelian tradition. Again, it seemed at one time to be generally agreed – and accepted by psychological investigators of decision making (see Slovic et al. 1977) – that the rational way to base action on estimates of chance was to follow the rule: "Rank possible courses of action according to their conditional subjective estimations of utility." But this rule has been seriously challenged in recent years because it seems not to take proper account of the difference between actions as symptoms, and actions as causes, of states of affairs that we act to promote or avert (Jeffrey 1980).

Great care has certainly to be taken also in selecting the normative criteria by which the correctness of subjects' probability judgments is assessed. In one experiment, for example, subjects were told that in a certain town blue and green cabs operate in a ratio of 85 to 15, respectively. 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. The subjects were then asked: what is the probability (expressed as a percentage) that the cab involved in the accident was blue? The median estimated probability was .2, and investigators (Kahneman & Tversky 1972a) claim that this shows the prevalence of serious error, because it implies a failure to take base rates (that is, prior probabilities) into account. Kahneman and Tversky commented: "Much as we would like to, we have no reason to believe that the typical juror does not evaluate evidence in this fashion." Lyon and Slovic (1976) have confirmed the robustness of the phenomenon, which is impervious to variations in the topic, numerical details, and sequential formulation of the story told to the subjects (with the proviso that blue and green cabs were present in equal numbers during the tests on the witness). And they complain that "since the world operates according to Bayes's theorem, experience should confirm the importance of base rates" despite the apparent failure of subjects to recognize that it does so.

At best, these experiments would constitute a test of their subjects' intelligence or education, since the ordinary person might no more be expected to generate Bayes's theorem spontaneously than Bernoulli's. But in fact it is doubtful whether the subjects have made any kind of mathematical error at all. The experimenters seem to be reasoning as follows. In the long run, they say to themselves, the witness may be expected to make 68% correct identifications of a cab as blue ($\frac{4}{5} \times 85\%$), 3% incorrect identifications of a cab as blue ($\frac{1}{5} \times 15\%$), 12% correct identifications of a cab as green ($\frac{4}{5} \times 15\%$), and 17% incorrect identifications as green ($\frac{1}{5} \times 85\%$). Therefore he will altogether make 29% identifications as green, and the fraction of them that will be incorrect is $\frac{17}{29}$. Consequently, according to the way in which the experimenters seem to be reasoning, the probability that the cab involved in the accident was blue is $\frac{17}{29}$, not $\frac{1}{5}$.

But this last step is a questionable one. The ratio $\frac{17}{29}$ is the value of the conditional probability that a cab-colour identification by the witness is incorrect, on the condition that it is an identification as green. Jurors, however, or people thinking of themselves as jurors, ought not to rely on that probability if they can avoid doing so, since reliance on it assumes the issue before the court to concern a long run of cab-colour identification problems – whereas in fact it concerns just one problem of this type. Jurors here are occupied, strictly speaking, just with the probability that the cab actually involved in the accident was blue, on the condition that the witness said it was green. And the latter probability is equivalent in the circumstances to the probability that a statement to the effect that the cab actually involved in the accident was green, is false, on the condition that the statement is made by the witness. If the jurors know that only 20% of the witness's statements about cab colours are false, they rightly estimate the probability at issue as $\frac{1}{5}$, without any transgression of Bayes's law. The fact that cab colours actually vary according to an $\frac{85}{15}$ ratio is strictly irrelevant to this estimate, because it neither raises nor lowers the probability of a specific cab-colour identification being

correct on the condition that it is an identification by the witness. A probability that holds uniformly for each of a class of events because it is based on causal properties, such as the physiology of vision, cannot be altered by facts, such as chance distributions, that have no causal efficacy in the individual events. For example, if the green cab company suddenly increased the size of its fleet relative to that of the blue company, the accuracy of the witness's vision would not be affected, and the credibility of his testimony would therefore remain precisely the same in any particular case of the relevant kind.

The same point can be put another way by emphasising the difference between probability functions that measure relative frequencies and probability functions that measure causal propensities (see part I, section 2). Propensity-type probabilities may be *estimated from* frequencies in appropriate samples (as with the witness's reliability), but what is actually evaluated is something different: a propensity, not a frequency. And propensity-type probabilities can be derived for individual events because they are predictable distributively. So it is natural to suppose that this is the kind of probability with which a jury is properly concerned, whereas the mere relative frequency of blue and green cabs is an accidentally accumulated characteristic of the town's cab population, considered collectively, and does not generate any causal propensity for the particular cab in the accident. Of course, if no testimony is mentioned and subjects know nothing except the relative frequency of the differently coloured cabs, then no causal propensity is at issue and the only basis for estimating the required probability is indeed the relative frequency. And this is in fact the kind of estimate that the investigators have then found to occur under experimental conditions (Lyon & Slovic 1976, p. 294).

The issue here is an important one since it has many ramifications. If the investigators had been right to impugn the rationality of commonsense judgments in the above example, it would have certainly been difficult to defend the continued use of lay juries. Consider too what you yourself would decide in the following circumstances. You are suffering from a disease that, according to your manifest symptoms, is either A or B. For a variety of demographic reasons disease A happens to be nineteen times as common as B. The two diseases are equally fatal if untreated, but it is dangerous to combine the respectively appropriate treatments. Your physician orders a certain test which, through the operation of a fairly well understood causal process, always gives a unique diagnosis in such cases, and this diagnosis has been tried out on equal numbers of A- and B-patients and is known to be correct on 80% of those occasions. The tests report that you are suffering from disease B. Should you nevertheless opt for the treatment appropriate to A, on the supposition (reached by calculating as the experimenters did) that the probability of your suffering from A is $^{19}/_{23}$? Or should you opt for the treatment appropriate to B, on the supposition (reached by calculating as the subjects did) that the probability of your suffering from B is $^{4}/_{5}$? It is the former option that would be the irrational one for you, qua patient, not the latter; and in a rather comparable experimental situation (Hammerton 1973) subjects tended in fact to judge the matter along just those lines. Indeed, on the other view, which is the one espoused in the literature, it would be a waste of time and money even to carry out the tests, since whatever *their* results, the base rates would still compel a more than $^{4}/_{5}$ probability in favour of disease A. So the literature under criticism is propagating an analysis that could increase the number of deaths from a rare disease of this kind.

Admittedly, the standard statistical method would be to take the prior frequency into account here, and this would be absolutely right if what was wanted was a probability for any patient considered not as a concrete particular person, not even as a randomly selected particular person, but simply as an instance of a long run of patients. The administrator who wants to secure a high rate of diagnostic success for his hospital at minimal cost would be right to seek to maximise just that probability, and therefore to dispense altogether with the tests. But a patient is concerned with success in his own particular case, not with stochastic success for the system. So he needs to evaluate a propensity-type probability, not a frequency-type one, and the standard statistical method would then be inappropriate. Note, however, that the causal propensity analysis does not involve any repudiation of Bayes's theorem. It is just that the prior probabilities have to be appropriate ones, and there is no information about you personally that establishes a greater predisposition in your case to disease A than to disease B. We have to suppose equal predispositions here, unless told that the probability of A is greater (or less) than that of B among people who share all your relevant characteristics, such as age, medical history, blood group, and so on. An analogous supposition has to be made about the cab colours, unless we are told that because of faulty maintenance, say, the probability of a blue cab's being involved in accidents that share all the relevant characteristics of the present one, such as poor braking, worn tires, and the like, is greater (or less) than that of a green cab's being involved. Similarly, in a criminal law court the object is to do justice in each individual case, without taking a defendant's past criminal record, if he has one, into account. But it is easy enough to imagine analogous cases in which a shoplifter, say, would escape conviction on the basis of probabilistic testimony about identification, if the relative frequency of honest shoppers could be cited in his defence! Or consider an exmple very like that cited by Todhunter (1949/1865, p. 400) in connection with the danger of applying the standard statistical method – which he traces to Condorcet – indiscriminately. A witness of 99.9% reliability asserts that the number of the single ticket drawn in a lottery of 10,000 tickets was, say, 297: ought we really to reject that proposition just because of the size of the lottery?

The difference between frequency probability and propensity probability is a difference between two functions that both satisfy the formal axioms of the classical calculus of chance. The two functions differ in their semantics, that is, with regard to the nature x and y must have, and the relation they must bear to one

another, when, for a particular n, it is true that $p(x \mid y) = n$. But both functions have the same logical syntax; that is, each satisfies a multiplicational law for conjunction, a complementational law for negation, and so on. Nevertheless (as remarked in part I, section 2) it should by no means be taken for granted that all valid types of probability judgment in everyday reasoning can be modeled by functions that share this syntax.

For example, it has been held (Kahneman & Tversky 1972b; 1973; 1974; Tversky & Kahneman 1974) that intuitive judgments of probability are biased towards predicting that outcomes will be similar to the evidence afforded by typical cases. It is claimed that people use a representativeness heuristic as a rough-and-ready, though often misleading, guide in their probabilistic reasoning. But the validity of this claim depends on the assumption that such a judgment about degree of representativeness has to be interpreted as a means towards drawing some conclusion about probability in a sense of that term that conforms to the classical calculus of chance. If instead we abandon that assumption, we can avoid imputing any fallacies here. We can suppose that the judgment of representativeness leads to a conclusion about probability in a sense in which an inference from representativeness to probability is always quite legitimate – albeit a sense that conforms to principles different from those derivable within the calculus of chance. In fact, these principles can be shown to be implicit in the logic of controlled experiment, which was first developed by Francis Bacon (Cohen 1979). Bacon, in the preface to his *Novum Organum*, described the central concern of his own enquiry in just the same terms as Bernoulli (1713, p. 211) described his, namely, the determination of "degrees of certainty." But Bacon's method defines a different concept of probability from Bernoulli's (Cohen 1980b). Hume (1739) called it "probability arising from analogy," and he wrote:

> Without some degree of resemblance, as well as union, 'tis impossible there can be any reasoning; but as this resemblance admits of many different degrees, the reasoning becomes proportionally more or less firm and certain. An experiment loses of its force, when transfer'd to instances, which are not exactly resembling; tho' 'tis evident it may still retain as much as may be the foundation of probability, as long as there is any resemblance remaining.

When all this is made precise and its implications are developed systematically, one can show that, in appropriate contexts, concern with representativeness is not a potentially fallacious heuristic but rather a quite reliable, albeit somewhat crude, mode of commonsense reasoning under conditions of uncertainty (Cohen 1979; 1980d). It appears otherwise only if evaluated against a type of normative theory that is inappropriate in the circumstances (though admirably appropriate in many other circumstances).

Conclusion

The upshot of all this may be summarised as follows. No doubt ordinary people often err in their reasoning, and such a mistake begins to be of scientific interest when it can be shown to instantiate some regular pattern of performance error. However, nothing in the existing literature on cognitive reasoning, or in any possible future results of human experimental enquiry, could have bleak implications for human rationality, in the sense of implications that establish a faulty competence. At best, experimenters in this area may hope to discover revealing patterns of illusion. Often they will only be testing subjects' intelligence or education. At worst they risk imputing fallacies where none exist.

ACKNOWLEDGMENTS

I am grateful for helpful comments on an earlier draft of this paper to Jonathan Adler, Gillian Cohen, the ten *BBS* referees, and participants in discussions at a meeting of the Society for Philosophy and Psychology at Ann Arbor (March 16, 1980) and at seminars at Johns Hopkins University (March 20, 1980), Witwatersrand University (July 23, 1980), Australian National University (August 20, 1980), Melbourne University (September 5, 1980), Queensland University (September 17, 1980), and Victoria University of Wellington (September 23, 1980). I am also grateful for valuable research assistance provided by the Research School of Social Sciences, Australian National University.

NOTES

1. The same is true for intuitions of grammaticalness, pace Sampson (1975).

2. This issue is too complex to be treated adequately here; for a useful review see Haack (1974).

3. I leave open here the much discussed question whether Lewis and Langford, 1959, Anderson and Belnap 1974, or some other system provides a better fitting logic of everyday reasoning.

4. For example, the intuition that *B* is deducible from *A* whenever *A*-and-not-*B* is inconsistent (Lewis & Langford 1959) clashes with the intuition that one may not deduce every proposition from an inconsistent one (Anderson & Belnap 1974). So, though consistency is normally an overriding ideal for theory construction, one cannot treat the demand for it as the only foundation needed for a theory of deducibility: other intuitions, too, have to be taken into account.

5. Cf. how, in a maximally specific case, the systematic model of a clinician's judgmental strategies may be a better predictor than the clinician's own judgment (Goldberg 1970).

6. The analogy with perceptual illusion (such as the Müller-Lyer) was also drawn by Chapman and Chapman (1967, p. 194) in their interpretation of the partly experimental and partly real-life data about erroneous use of Draw-a-Person tests in psychiatric diagnosis. Both here and in their work on the psychodiagnostic use of Rorschach cards (1969) they traced the source of illusory correlations to a powerful bias by verbal association, since subjects with no clinical experience at all tended to make the same erroneous correlations as many clinicians.

7. Apparently none of Wason's subjects objected, as would have been justified, that no finite number of questions and answers, whether falsificatory or verificatory, could prove such a hypothesis correct.

8. I take Grice to have established the mental or social reality of some such rules. In the logical context, however, he does not use them, as I do, to explain the alleged prevalence of the fallacy of illicit conversion. Instead he tries to use them to explain away the apparent inappropriateness of a truth-functional logic for the analysis of deductive reasoning in a

natural language, and in this he attempts an impossible task (see Cohen 1971; 1977a).

9. This has now been acknowledged by its authors (Kahneman & Tversky 1979).

Open Peer Commentary

Commentaries submitted by the qualified professional readership of this journal will be considered for publication in a later issue as Continuing Commentary on this article. Among the commentaries to appear in a forthcoming Continuing Commentary section are those by N. Daniels & G.E. Smith; R. Grandy; G.S. Kahn & L.J. Rips; C.R. Mynatt, R.D. Tweney & M.E. Doherty; R. Revlin; and G. Shafer. Time and length constraints prevented these commentaries from appearing with the target article, but the author (L.J. Cohen) has had the opportunity to see and to reply to them in his Response. Forthcoming commentaries are indicated by an asterisk (°) in the Author's Response.

Rational animal?

Simon Blackburn

Pembroke College, Oxford OX1 1DW, England

Jonathan Cohen's central contention sounds very radical: "nothing in the existing literature on cognitive reasoning, or in any possible future results of human experimental enquiry, could have bleak implications for human rationality, in the sense of implications that establish a faulty competence." But before researchers into our cognitive defects lay down their tools, they will want to reflect that this optimism is only protected by a "competence/performance" distinction: In Cohen's own terms it is quite possible that we suffer from systematic tendencies to poor performance, even if these can only be located within a general view of the subject that regards him as competent in the use of the kinds of argument that he has performed badly. But I doubt whether the argument Cohen offers supports even this dilute optimism. Its essence is that if we are to attribute faulty performance to an individual (e.g. illicit conversion of a conditional) we must be sure that the sentences he is using (e.g. English ones with "if then" syntax) are used to express conditionals in the sense of our logical theory. But doing this is in part discovering that the sentences *are* used with the correct deductive liaisons of conditionals. If the speakers didn't do this much, then we couldn't be sure that conversion was illicit at all (e.g. if the conditional form actually expressed belief in the biconditional).

Certainly it behooves us to look out for this possibility. But sometimes people might exhibit enough of a tendency to accept and reject certain inferences that this would put quite definite constraints on what they *could* be doing in using the "if then" construction; they might then proceed to judge and perform in such a way that *no* consistent model of their reasoning could be reconstructed. In that case there would be no defence of them along the lines that perhaps we mistranslate the connective. For example, we all incline to accept argument forms such as conditional proof; these are only valid if "if then" expresses material implication; we then disallow other properties of that operation. There is no consistent logic we *could* be obeying. If these things are true of us, then they point to a flaw in competence, not performance. Of course, as Cohen rightly says, we must be very careful to ascertain that people are actually doing such things. We are very quick to include collateral information in our processes of argument, which therefore look invalid to anyone who only takes account of the strict and literal content of our sentences. But that points to an experimental pitfall, not to an a priori certainty in our competence.

One might reply: Well, what if we are just vague in what we mean by "if then," or let it move amongst possible meanings? We would then preserve our rationality even in the face of the worst kind of result. We would; but only at the cost of a different kind of irrationality. A feature of our cognitive lives is irrational if it increases the likelihood of arriving at false judgment. A language that allowed systematic jumps in the meaning of logical particles would increase the likelihood of arriving at false judgment, unless the jumps were clearly signalled, so that people knew where they were. Since we clearly don't have such signalling ("now I'm going to use "if then" to mean "if and only if"), it would be irrational to use such a language.

When the witness has reported the green colour of the cab, we have a piece of evidence putting us in an "epistemic situation" which has the following relation to the truth: If in that situation we believed the cab to have been green, we would be right in 12 cases out of 23. In the diagnosis case, believing the result of the test, we would be right 4 times out of 23. It is bad to form such beliefs in such epistemic situations: It is doing something that increases the likelihood of arriving at false judgment. How did Cohen come to recommend it? As John Mackie [q.v.] rightly points out, the principles of the cases have been known since Hume, roughly, and Thomas Bayes, precisely.

I suspect that the root trouble is the metaphysical notion of a propensity, thought of as a particular real, but gradable feature of individual trials on a chance setup. We can protect ourselves from the harmful consequences of this notion only by invoking a battery of distinctions, as Mackie does: the patient's "abstract propensity" to contract a disease, versus his "propensity in the actual situation," versus the frequency with which patients get the disease, versus the epistemic probability that he will get it. This is surely too cumbersome to be helpful in guarding against the fundamental opportunity for error which the concept brings: the belief that "mere" statistical evidence is irrelevant to *me*, because my propensity to get the disease is what it is, regardless of other people's no doubt different propensities. I have argued elsewhere that legitimate talk about propensities is best seen as a projection of our concern in getting the best judgment we can about the individual case: It is not something to set over and against rational judgment, but our way of projecting the norms that guide rational judgment (Blackburn 1980). On my picture we may legitimately ask whether my propensity is different from the 4/23 chance that I have the B disease, but only because we can envisage a different, narrower reference class into which, ideally, I should be put. While there is no such class, to have any other than 4/23 confidence that I have the B disease is doing something that increases the likelihood of false judgment. [This would be apparent from the disastrous results of a hospital that institutionalized the tendency.] It is quite true that to get the single-case confidence from the long-run truth requires the exercise of what I have called the Population Indifference Principle: If the best epistemic situation you can get into about an individual case is one in which judging it to be X would be right $P\%$ of the time, then you should have $P\%$ confidence in that judgment. But this is constitutive of rational response to evidence. (Blackburn 1973).

When should we attribute a defect that increases the likelihood of false judgment to irrationality, and when to simple lack of education? The only illuminating suggestion I

can think of has a pleasant connexion with some of the experimental results. We would censure someone for possessing such a defect the more if ordinary life afforded him opportunities for self-correction (self-education) but he did not profit from the lesson. Thus, ordinary people who are not surveyors have no opportunity for practicing Pythagoras's theorem, and require teaching. But the experimental results seem to support the view that we show defects in our pursuit of truth mostly where we don't often pursue it, which is exactly where ordinary life has given us no opportunity to practice self-correction. Perhaps basic optimism about rationality where we need it is therefore justified. Still, if a jury finds me guilty, or a politician blows me up, it is small consolation to think that their mistake was due only to an unfortunate lack of practice in extending a basic competence outside their familiar fields.

Status of the rationality assumption in psychology

Marvin S. Cohen

Decision Science Consortium, Inc. Falls Church, Va. 22043

The claim that something cannot in principle be established experimentally does not itself, presumably, rest on experimental evidence. Thus, Professor Cohen appears to be offering a constitutive principle for the scientific study of ordinary human reasoning: It can never be regarded as systematically flawed. Errors, when they occur, must be attributed not to basic competence, but to features of performance such as fatigue, motivation, inattention, or lapse of memory. In this claim, however, the argument takes on the look of an empirical hypothesis. This ambiguity bears closer scrutiny. Unraveling its implications will reveal, I believe, that Cohen's thesis is neither plausible as an empirical hypothesis nor fruitful as a constitutive principle.

Consider the class J of all instances of "human reasoning." In evaluating human rationality, according to Cohen, one must use criteria based on the normative theory (or combination of theories) that best accounts for actual instances of reasoning; that is, that comes closest to generating the class J. Ignoring the issue of uniqueness, let us refer to the theory (or combination of theories) that is most successful in this sense as T. Now most instances of reasoning in J will turn out to be correct according to T (the class A), but others will not ($\overline{A}$). To explain $\overline{A}$, we introduce the class B, containing all instances in which performance factors cause a change in the outcome of reasoning. Now the claim with which we are concerned can be formulated thus: All members of $\overline{A}$ are members of B.

An opposing view maintains that errors in reasoning may be caused by the *same* processes that lead on other occasions to "correct" reasoning, and that no interference or distortion from other sources is necessary. Such processes are, of course, "heuristics": Even when properly used, they need not always produce the desired result.

The issue as stated seems amenable to experimental test, at least in principle. For example, fatigue, motivation, demands on attention, and memory can all be manipulated by standard experimental techniques, and the effect on a concurrent reasoning task can be observed. The question of interest would be whether evidence indicates that the use of heuristics increases with the disturbing factor, independently of any increase in errors. If it does, we have experimental support for the claim that heuristics replace more normatively appropriate procedures under conditions of stress. If the use of heuristics does not vary as a function of stress, on the other hand, it is likely that they are employed in ordinary unstressful human reasoning, with occasional erroneous consequences.

Let us imagine it established that some members of $\overline{A}$ are not members of $\overline{B}$. Now what can Cohen's reaction to this hypothetical state of affairs be?

1. He can reject the claim that all members of $\overline{A}$ belong to B, as a disconfirmed empirical hypothesis. But this would be an admission of human irrationality on experimental grounds, contrary to the main thesis of his paper.

2. He can conclude that T is not, after all, the best normative theory in accounting for J. Some different theory T' must be found, such that all instances of reasoning that are erroneous according to T' are in fact members of B. In effect, empirical data on performance determine the portion of J ($A' = J \wedge \overline{B}$) that must be generated by the selected normative theory.

3. As a final option, Cohen might retain T as the best normative account of J, but conclude that B does not correctly represent the role of performance factors. In its place we put B', which is simply *defined* as $\overline{A}$, the class of all erroneous instances of reasoning. This is the opposite tack from (2), where the empirical theory of performance determined the domain of the normative theory. Here instead the normative theory defines the theory of performance.

Unfortunately, options (2) and (3) may have quite unpalatable consequences. For concreteness, consider the availability heuristic. An appropriate normative theory, proposed by Cohen for this situation, features the assumption (T) that frequency in a population can be estimated by frequency in a representative sample. Errors occur, he says, only when performance factors distract subjects from determining that the available sample is, in fact, representative. Now it is not impossible that experimentation would disconfirm this, showing that no identifiable performance factors are responsible for the observed results. Is there a revised normative theory, T′ that, as in option (2), depicts the observed behavior as rational? The critical point is this: There is no guarantee that such a normative theory, T′, exists.

Such a theory is guaranteed only if "normative" is construed loosely enough. We simply adopt as T′ the "theory" corresponding to the heuristic itself: that frequency is availability, or alternatively, that all available samples are representative. This, of course, trivializes Cohen's thesis, and he will have none of it.

The result of applying option (3) is analogous. The theory of performance, defined by reference to T, is vacuous. We are at liberty to invoke "inattention," "lapses of memory," and "failures of motivation" whenever and wherever we need to in order to preserve T as a theory of competence. But these terms have left behind them any connection to their use in an experimental science.

What can we conclude about the status of Cohen's thesis that all errors are attributable to performance? If it is an empirical hypothesis, we must be prepared to relinquish it in the event that (a) performance factors are not experimentally supported for some instances of $\overline{A}$, and (b) no plausible alternative normative theory is available to account for $J \wedge \overline{B}$. Nor should these eventualities be regarded as improbable. Ethology reveals numerous instances of behaviors that are adaptive in certain contexts but not in others. Evolutionary success does not imply that a structure or trait has been appropriate wherever it occurred, only that it has been appropriate often enough.

As a constitutive principle for the psychological study of human reasoning, Cohen's thesis seems off the mark. It can lead, *in principle*, either to the adoption of utter nonsense as a normative theory, or else to the abandonment of the empirical study of attention, memory, and motivation.

Whether *in fact* it does so, depends on the outcomes of experiments. It *might* turn out that competence, as defined by normative theory T, is the exact complement of performance, as defined by experimental psychology. But this result would be quite fortuitous in light of the evolutionary considerations referred to in the previous paragraph. And the mere possibility of a negative outcome is enough to warrant rejection of Cohen's principle.

Where the thesis does have plausibility, perhaps, is as a "constitutive principle" for ordinary reasoning about our fellow human beings. In this sense, it is akin to the principle of charity in translation: We attempt to construe another's words and actions so as to make them seem as rational as possible. If someone denies the obvious, we try to construe his words in a different sense. Ad hoc "lapses of attention" are invoked where we fail. It is in this sense only that human reasoning ability is to be regarded as "flawless."

Empirical psychology is not evaluative in this way. Heuristics that correctly describe reasoning will occasionally violate obvious norms and thus not be sensibly interpretable as "beliefs" of the subject. For the same reason, such heuristics are not part of a theory T that purports to be normative. Such constraints can only hinder the effort to find an account of reasoning that is consistent with theories of attention, memory, and motivation – however flawed our reasoning ability turns out to be.

The persistence of cognitive illusions

Persi Diaconis[a] and David Freedman[b]

[a]*Statistics Department, Stanford University, Stanford, Cal. 94305 and* [b]*Statistics Department, University of California, Berkeley, Cal. 94720*

Introduction. Cohen asserts that "normative theory . . . is itself acceptable . . . only so far as it accords, at crucial points, with the evidence of untutored intuition." He goes even further: "Ordinary human reasoning . . . sets its own standards." However, our view is that untutored intuition is often wrong, and much can be learned by examining common mistakes in reasoning.

A focal point for our discussion will be the idea of "cognitive illusions." Cohen introduces this idea as follows: "subjects are induced to indulge in a form of reasoning that on a few moments' prompted reflection they would be willing to admit is invalid." Much of the research under review is said by Cohen to demonstrate at most the existence of such cognitive illusions. On our view, many of the illusions cannot be dispelled by a "few moments' prompted reflection," or several months of college teaching; if dispelled, the illusions seem to return in full force the next time a similar situation comes along. Such illusions seem to be rooted very deeply in the human mind. Our main point is that these illusions exist and are worth studying. We will draw three examples from Cohen's article: More, we will try to show that Cohen himself commits all three fallacies. The first does not seem to have been named: We dub it "the fallacy of the transposed conditional."

The fallacy of the transposed conditional. In part II, section 4, Cohen discusses the example of a witness to a lottery; the "reliability" of this witness is given as 99.9%. There are 10,000 tickets in the lottery. One is drawn at random; the witness says it was 297. What is the probability that it really was 297? According to Cohen, the "standard statistical method" takes into account the number of tickets in the lottery; he considers this to be wrong. However, there is a real ambiguity in Cohen's statement of the problem. Let X be the number drawn in the lottery, and Y the number reported by the witness. Does "reliability" mean $P(X = n \mid Y = n)$ or $P(Y = n \mid X = n)$? If the former, contrary to Cohen's assertion, the number of tickets in the lottery will not affect the probability of error, as computed by the "standard statistical method." If the latter, the problem is incompletely specified: the missing information being $P(Y = n \mid X = m)$ for $n \neq m$.[1] As far as we can see, Cohen reaches his conclusion by sliding from one of these interpretations of reliability to the other. Confusing $P(A \mid B)$ with $P(B \mid A)$ is "the fallacy of the transposed conditional." It is the statistical analogue of confusing "p implies q" with "q implies p."[2] Both errors are made frequently, even by people who ought to know better. Cohen cites Todhunter (1949/1865) in his discussion of witness reliability, and Todhunter cites de Morgan (1856). De Morgan got it right, in 1856; Cohen's error is an example of a persistent cognitive illusion.

The base rate fallacy. Cohen construes the base rate fallacy (part II, section 4) as a "test of . . . intelligence or education." We do not agree. To see why, consider his discussion of the jury trial involving eyewitness identification of a taxicab. Was the cab blue or green? At one point Cohen concedes that, in the absence of other evidence, base rates are a "basis for estimating the required probability": there is an 85% chance for the cab to be blue and 15% to be green. Just before this, Cohen stipulates that, under the relevant conditions, there is an 80% chance that the witness can correctly distinguish between the two colors. In short, Cohen seems to accept the two hypotheses of Bayes's theorem.[3] This theorem, like any other, is a statement of the form "if p, then q." Stripped of irrelevant detail, Cohen's argument amounts to granting p but denying q. Indeed, as soon as the jury has both the base rate and the witness, Cohen instructs it to disregard the base rate. As he insists later, "we have to suppose equal predispositions here. . . ."[4] Cohen's argument for disregarding the base rate is as follows: "A probability that holds uniformly for each of a class of events because it is based on causal properties, such as the physiology of vision, cannot be altered by facts, such as chance distributions, that have no causal efficacy in the individual events." In this lapidary sentence, Cohen not only commits the base rate fallacy, but positively defends it. Evidently, the base rate fallacy is a persistent cognitive illusion.

The gambler's fallacy. According to Bernoulli's theorem, if a fair coin is tossed a large number of times, the proportion of heads will be close to .5, with high probability. The deviation between the actual number of heads and the expected number is swamped by the number of tosses.[5] There is no compensation: A run of heads is as likely to be followed by a head as by a tail. In short, the coin has no memory. There is good experimental evidence to support the theory. However, as Tversky and Kahneman (1974) show, and as we see over and over again in the classroom, people have a hard time believing the theory. The belief that a run of heads is likely to be followed by a tail is called "the gambler's fallacy." The fallacy is as current today as it was in Bernoulli's time, and as hard to extirpate. This is our third example of a persistent cognitive illusion.

Consider Cohen's discussion (part II, sections 3 and 4). He makes three points, which can be summarized (a little brutally) as follows:

i. People who believe that the coin has a memory should be interpreted as believing that the coin has a memory.
ii. There is in fact no evidence that people actually do believe the gambler's fallacy; such evidence is in principle impossible to obtain; and even if it could be obtained, it would have no psychological significance.
iii. Someone with infinite resources, who believed the fallacy, could make money by betting according to in-

tuition, if only casinos tossed fair coins and lifted the house limit on the size of bets.

Many gamblers would be very happy to meet the person described in (iii). Leaving that vision aside, Cohen's discussion obscures the two main points:

i. The gambler's fallacy is a fallacy.
ii. People believe it anyway.

Conclusion. Both of the present commentators make a living by teaching probability and statistics. Over and over again, we see students and colleagues (and ourselves) making certain kinds of mistakes. Even the same mistake may be repeated by the same person many times. Cohen is wrong in dismissing this as the result of "mathematical or scientific ignorance." Analysis of such repeated mistakes is crucial to successful teaching. The work of Tversky and Kahneman (1974) is very helpful in this respect. Their psychological explanations are, at a minimum, valuable heuristics which help organize experience and lead to new understanding. Their examples are wonderfully illuminating.

By comparison, Cohen's analysis seems to be both sterile and wrong. There is nothing in his target article that helps us to understand the difficulties people have in dealing with standard probability theory. In other publications, he has attempted to develop some alternatives to the standard theory. But his alternatives are cranky. Standard theory says some useful things about the empirical world, but the insights often contradict basic intuition. As a result, people have a hard time learning the theory. Cohen's solution seems to be to revise the theory so it will accord with his a priori idea of "untutored intuition." This must be the philosopher's version of throwing out the baby with the bathwater.

NOTES

1. On the first interpretation of reliability, $P(X = n \mid Y = n) = .999$ by assumption, regardless of the number of tickets in the lottery. Now take the second interpretation. There are 10,000 tickets in the lottery, so X is uniform from 1 to 10,000. By Bayes's theorem, $P(X = n \mid Y = n) = a/a + b$: where $a = P(Y = n \mid X = n) \cdot P(X = n) = .999 \times 1/10{,}000$; and b is the sum over $m \neq n$ of $P(Y = n \mid X = m) \cdot P(X = m)$, so b is between 0 and $(1 - .999) \times 9{,}999/10{,}000$. Thus, $P(X = n \mid Y = n)$ is between .09 and 1, depending on the values of the missing parameters. That is quite a range. If, for example, $P(Y = n \mid X = m$ and $Y \neq m)$ is 1/9,999, then the number of tickets in the lottery really does enter into $P(X = n \mid Y = n)$, both in theory and in practice. On these assumptions, an observer who uses Bayes's rule would have a better chance of being right than an observer who uses Cohen's calculus.

2. $P(A \mid B)$ can be interpreted as the probability of A, if you know that B has occurred.

3. These hypotheses may be open to question, but Cohen does not raise such issues. And we doubt that these issues are relevant to an understanding of the base rate fallacy.

4. This must be an application of Laplace's principle of insufficient reason, first published around 1780: another example of a long-lived cognitive illusion.

5. This deviation is likely to be on the order of .5 times the square root of the number of tosses (DeMoivre's theorem).

Rationality and the sanctity of competence

Hillel J. Einhorn and Robin M. Hogarth

Center for Decision Research, Graduate School of Business, University of Chicago, Chicago, Ill. 60637

We are sympathetic to a number of points made by Cohen: the importance of context in assessing error in judgment; the crucial role of intuition in judging rationality; the conditional nature of normative models, and so on. Indeed, these issues are discussed at length in our review of behavioral decision theory (Einhorn & Hogarth 1981). What is new in Cohen's paper is the proposed distinction between competence and performance in reasoning tasks. This distinction not only provides the basis for many of his criticisms of the existing literature, it also serves to justify the negative answer to the question posed in his title. We consider this distinction in detail below but note that Cohen's basic arguments about human rationality fall into the class of explanations termed "imaginary reconstructions" (Lewontin 1979), that is, attempts to rescue rationality in behavior despite much evidence to the contrary. While we believe that some psychologists have gone too far in describing people as "cognitive cripples," Cohen has overreacted and gone to the other extreme. To show this, we first consider the competence-performance distinction and its implications.

Consider that performance on some inference task is a joint function of competence, knowledge, motivation, and environmental factors (including random variation). This can be expressed by the conceptual equation

$$\text{Performance} = f\,(\text{Competence, Knowledge, Motivation, Environment}) \quad (1)$$

If, as seems likely, these factors trade off to some extent in producing a particular level of performance (e.g., f could be multiplicative or additive), there is an immediate implication: Variability in performance can be explained by changes in factors other than competence. Thus, if rationality is to be defined in terms of a constant level of competence, we should have no difficulty in attributing any defect in performance to one or more of the other factors. Thus, Cohen's position is untestable. Moreover, we are puzzled by the title of his target article, since the above position makes it impossible to demonstrate that *any* behavior, irrespective of whether or not it is experimentally produced, is irrational. Perhaps we have oversimplified Cohen's position, and a more complete analysis is needed. In particular, how are the various factors in (1) defined and used in the paper?

Consider what Cohen means by competence. He states: "To ascribe a cognitive competence . . . is to characterise the content of a culturally or genetically inherited ability which, under ideal conditions, every member of the community would exercise in appropriate circumstances." Later on, in discussing information-processing mechanisms he says, "The structure or design of such a mechanism must account for the relevant competence." This seems to suggest that our information-processing equipment, which is our biological heritage from a long evolutionary process, is sacrosanct and not to be considered defective in any way. Furthermore, although Cohen agrees that human rationality is bounded, he states that this is the case "in practice," thus implying that limited short-term memory, selective attention, lack of attentional control, and so on, are not inherent biological limitations in design but rather performance defects caused by insufficient motivation, lack of knowledge, or environmental effects. We call this type of competence "design competence" and consider it shortly. However, Cohen seems to have another meaning of competence in mind when he includes "culturally inherited ability" as part of his definition. This seems to imply a "competence for learning" as well. Indeed, in his example of a child being inferior in competence to an adult, the result can be seen as due to either a lack of maturation in the biological structures used in reasoning (design competence) or the lack of learning about and experience in performing certain mental operations (competence for learning). [See also Brainerd: "The Stage Question in Cognitive-Developmental Theory" *BBS* 1(2) 1978.]

If one asks for the rationale underlying the sanctity of design competence, or, to put it another way, why we *must* assume an underlying rationality for structure/design, no answer other than a naive evolutionary adaptationist position

is given. This position is most clearly articulated in an earlier paper (Cohen 1980b) in which we are told (p. 91) that the continued use of fallacious reasoning would have dire, and thus highly unlikely, genetic implications: "How strange that the top species in an evolutionary struggle for survival should end up with such lethal genes!" Whatever the merits of talking about the lethality of genes, we do not find the existence of persistent dysfunctional mechanisms to be incompatible with an evolutionary framework. This is particularly the case when one realizes that humans also adapt the environment to overcome their deficiencies (invent eyeglasses to correct poor eyesight; develop insulin for diabetics, and so on). Moreover, the time frame of evolution is such that we don't know whether any particular mechanism is being selected for or against, or is a vestige without function (Eichorn & Hogarth 1981, pp. 58–59; also see Lewontin 1978). Thus, Cohen's claims for design competence rest on no grounds that we can discern.

A second problem for Cohen's analysis is that a distinction between the competence for learning and *what* is learned is unclear. For example, one can agree with Cohen that lay adults who have not studied probability theory should not be faulted for failing to respond as though they had. However, how far should this argument be taken? Should one consider ignorance of formal physics a valid excuse for an adult who attempts to fly by jumping off a roof? Since people experience both sampling variability and the forces of gravity, at what point is it necessary that they receive explicit instruction in the laws of nature when judging the rationality of acts? While we have no answer, this is a more complicated question than Cohen seems to realize. In fact, the general difficulties involved in correctly learning about the quality of one's judgments and choices are ignored. While it could be argued that much of the literature Cohen criticizes does the same, the fact that learning and adaptation play a certain role in defining rationality (cf. Simon 1978) makes this a serious omission. For example, numerous investigators have discussed the difficulties of learning from outcome feedback, especially in probabilistic environments (e.g., Brehmer 1980; Einhorn & Hogarth 1978; Hammond 1978). Moreover, the disappearance of dysfunctional behavior is not assured by feedback and may even be strengthened by it (Einhorn 1980). In addition, there is evidence of biological limitations in the learning process itself (Seligman 1970). Thus, as was the case for design competence, we can find little support for an inherent rationality in the competence for learning. [See also Johnston: "Contrasting Approaches to a Theory of Learning" *BBS* 4(1) 1981.]

Where does this critique of Cohen leave us with respect to defining an error of reasoning in particular and the rationality of thought in general? First, Cohen's definition, whereby "nothing can count as an error of reasoning among our fellow adults unless even the author of the error would, under ideal conditions, agree that it is an error," is too vague to be useful (what, for example, are "ideal conditions"?) and conceptually unsatisfactory. It is not difficult, for instance, to think of situations in which real errors are not acknowledged as such (as in rationalizations) or errors are admitted where none has occurred (as in Galileo's retraction). Unlike those working in the field of perceptual illusions, we do not have commonly accepted yardsticks by which to assess mistakes. Instead, we must rely on the very judgment we wish to measure. Unfortunately, the elements that make up such judgments remain elusive.

With regard to the general issue of the rationality of behavior, we have given our own views on this matter elsewhere (Einhorn & Hogarth 1981). Briefly, judgments of rationality involve a mixture of the efficiency by which means secure ends (instrumental rationality – see Tribe 1973) and the goodness of the ends themselves (moral rationality). Thus, a calculating mass murderer may be judged as irrational to the degree that the attainment of despicable goals is accomplished with efficiency. Moreover, even internal criteria for rationality, such as consistency and coherence of behavior, are not without difficulty. For example, paranoids exhibit remarkable consistency and coherence in their belief systems. However, it is extremely difficult to convince such people that they have made errors in reasoning. [See Colby: "Modeling a Paranoid Mind" *BBS* 4(4) 1981.]

Although this comment has focused on our disagreement with Cohen, we do in fact agree with many of his opinions. Indeed, the need to study reasoning in natural environments seems to us to be especially important (Brunswik 1956). We cannot assume, as do many experimental psychologists, that performance in simple environments can necessarily be generalized to situations rich in a diversity of informational cues, where the use of multiple cues of equivocal validity is the rule rather than the exception (Hammond 1955; Postman & Tolman 1959). To quote Toda: "Man and rat are both incredibly stupid in an experimental room. On the other hand, psychology has paid little attention to the things they do in their normal habitat; man drives a car, plays complicated games, and organizes society, and rat is troublesomely cunning in the kitchen" (1962, p. 165).

On defining rationality unreasonably

J. St. B. T. Evans and P. Pollard

Department of Psychology, Plymouth Polytechnic, Plymouth, PL4 8AA, England

Cohen has argued that human irrationality cannot be demonstrated by the "existing literature" or by "any possible future results of . . . experimental enquiry" into deductive and probabilistic inference, in that a faulty *competence* can never be established from observation of faulty performance. In so arguing he adopts a philosophical stance that is impossible to refute and is of little practical relevance to the scientific study of human reasoning.

Cohen copies Chomsky's treatment of language in two respects. He argues for a distinction between underlying competence and observed performance and also insists that the definition of what constitutes a rational inference – by analogy to what constitutes a grammatical sentence – is its intuitive acceptability to ordinary people. There are thus two grounds on which apparently irrational behaviour may be explained: Competence may be disguised by the operation of performance factors, or else an a priori normative definition of rationality may be disqualified.

In reasoning experiments, the tasks normally require subjects to make intuitive assessments of the validity of inferences. By Cohen's definition, their responses can only be counted as errors if "the author of the error would, under ideal conditions, agree that it is an error." In this event the subject is said to be suffering from a "cognitive illusion" or a lack of appropriate education. Here we encounter the first major problem with Cohen's position. No criteria are specified for the definition of "ideal conditions," so we can never actually know when the subject is being "rational" or when he is exhibiting an illusion. The subject's agreement about an error is, in fact, of no use in this context. Although the experimenter can explain a problem and cause a subject to admit an error, on many problems the experimenter could just as easily persuade a subject that a normatively correct response was incorrect. Similarly, experimental manipulation of the effect of an "illusion" can produce spuriously high frequencies of normatively correct responses.

This problem becomes apparent in Cohen's discussion of

the Wason (1966) selection task as an example of a cognitive illusion. Naive subjects make a logical error on this task which Cohen terms a failure to apply the law of contraposition. That subjects do in fact possess the competence to apply this rule, he argues, is shown by the fact that responses tend to conform to logical requirements when the problem is phrased in realistic rather than abstract terms. Cohen himself appears to be suffering from a cognitive illusion here. He has argued that the failure to observe a logical response in some circumstances cannot be taken to imply that the subject lacks competence for the inference, as he may be "led astray" by a cognitive illusion. However, Cohen then expects us to accept the notion that observing the inference in other circumstances *does* mean that the subject possesses competence for its execution. "Competence" may be as illusory as "incompetence."

Perhaps Cohen would wish to argue that the use of familiar materials reduces the chance of illusory behaviour. He states that "most people manage to apply their logical competence without ever formulating it expressly at a level of generality sufficient for it to be readily applicable to wholly unfamiliar tasks." He is on shaky ground here. First, Cohen's definition of competence becomes unclear if it is held to be situation specific. Second, it is not in general the case that more realistic material improves logical performance. For instance, a priori beliefs about, or emotional attitudes towards, the content of a reasoning problem can be a major cause of erroneous inference (see Revlin & Leirer 1978; and, for a review, Evans 1982).

Another performance factor proposed by Cohen relates to the subject's degree of appropriate education and training. It is not clear whether he would wish to pursue the analogy with Chomsky's linguistics. To do so would involve the assumption of a genetically predetermined competence system, which could be "released" by appropriate experience, but whose fundamental structure could not be altered. If he believes this, then it should be possible to formulate a specific competence theory with testable consequences. If not, then Cohen's "competence" refers simply to learning capacity. [See also Chomsky: "Rules and Representations" *BBS* 3(1) 1980.]

The argument about education reveals an important inconsistency in Cohen's position. Unlike ethical problems, with which Cohen seeks to compare them, deductive or statistical reasoning problems lead to solutions that can be empirically validated. For instance, the gambler's fallacy can be said to be wrong because, if a person afflicted by the fallacy bet at subjectively reported "fair odds," he would tend to lose money. Cohen appears to accept the importance of sample size, although most subjects do not intuitively do so. It thus emerges that there *is* a normative system (which, at least in certain cases, Cohen himself accepts) that does not accord with general intuition. Cohen accounts for this discrepancy by arguing that subjects cannot be expected to know, but have the competence to learn, the principles. However, it follows from this that the lay person does not have the competence to apply the principle to relevant problems. This is the type of competence that is of interest to the psychologist.

In our view, the question of subjects' ability to behave rationally if properly trained to do so, or if not afflicted by certain irrational "illusions," diverts attention from the main purpose of research into human inference. It is theoretically desirable to understand the factors that actually determine inferential behaviour, and practically important to consider the consequences for actions and decisions in real life contexts. For example, Nisbett and Ross (1980) claim that people are far more influenced by vivid than by pallid information, regardless of its objective predictive power. A very good example of this is provided by Cohen himself. In the case of the illness problem, Cohen is clearly affected, and is inviting the reader to be affected, by an availability bias. The salience to the individual of a personal test is very high and far more vivid than tabulations of population base rates. The fact that Cohen's argument would lead to 15 more deaths than necessary out of 23 people diagnosed as having illness B, attests to the importance of investigating the possible effects of such biases in real life contexts.

A bias towards vividness might well mean that a powerfully placed decision maker will act on the basis of unrepresentative, but highly vivid, personal experiences or anecdotes and ignore the "dull" results of large, well designed statistical surveys. It is of no practical value to consider whether such behaviour can, by some philosophical device, be deemed to be "rational." It is evident that such behaviour is *undesirable*, in the sense that it is likely to produce inefficient decisions and costly errors. The point is that performance factors are not a theoretical complication to be removed in a search for rational competence. Understanding performance is the very essence of the scientific study of human inference.

Can any statements about human behavior be empirically validated?

Baruch Fischoff

Decision Research, A Branch of Perceptronics, Eugene, Ore. 97401

Two hundred years ago, psychology and philosophy were largely inseparable. Since they separated, the twain have seldom met. Few psychologists receive any training in philosophy, follow developments in philosophy, or even explore the way philosophers develop their own research topics. For their part, few philosophers are trained in the experimental study of the mind; few have "hands-on" experience with the intricacies and frailties of such work, or even track the research literature relevant to their own pedagogical efforts. Sallies across this particular disciplinary boundary, like those across others, often involve deprecating attacks or simplistic "solutions" to the others' problems, advanced with missionary or demagogic disrespect. Such miscommunication may be worse than no communication at all.

Before one concludes that this state of affairs is a reflection of lamentable ignorance or arrogance, it may be worthwhile to ask what philosophers and psychologists could or should be learning from one another, at least entertaining the thought that they would be better off extending a warm smile and a deaf ear to one another. Although the topics of the two fields are ostensibly similar, their agendas may be sufficiently different that mutual attention is likely to lead to mutual confusion. Interdisciplinary borrowing always runs the risks that arise from adopting methods and messages without the appreciation of the cautions accompanying their use that is only acquired through apprenticeship in a discipline.

Thus, the philosophers' chosen "isolation" may be explained by a feeling that psychologists have a less rigorous notion of what it means to get to the bottom of things. Rather than polishing their concepts, psychologists seem to have an insatiable need to get on with their business of collecting and interpreting data. In doing so, they make auxiliary assumptions about the nature of people and reality that are accepted by their colleagues as a matter of faith and taste. Often, as research areas gain self-confidence, attention will turn to testing these assumptions – at the price of making other ones. Philosophers have identified the indeterminacy of induction; psychologists must live with it, running the risk that entire research fields will prove to be poorly moored.

The psychologist who does peek across the border may feel frustrated by philosophers' fascination with the conceivable, rather than the probable; by philosophers' interest in identifying the full range of assertions that might be made, rather

than in testing the subset that might reasonably be subscribed to; by philosophers' penchant for keeping the debate on fundamental issues alive at all costs, rather than reaching some tentative conclusions on subsidiary issues and then moving on to new ones. The psychologist may acknowledge the legitimacy of abstract discussions about what the mind might entail, but prefer the more clumsy process of poking it to see what it does.

In this light, a case might be made that psychologists and philosophers would do well to ignore one another. The putative results that each offers are grounded in pursuits that are somewhat irrelevant to the other; moreover, they hide subtle issues to which outsiders are insensitive. Perhaps a more realistic aspiration at the present stage of nonintegration is to learn to respect and gradually adopt some of one another's guiding predispositions. Perhaps at some future date, cooperative arrangements will be possible.

One predisposition that psychologists might learn from philosophers is to strive for maximum clarity in the terms one uses, so as to facilitate comparison between studies, avoid internal contradictions, and present coherent tasks to subjects. Another is to explore in depth the logical relationships between different theories and hypotheses, so as to reduce the incidence of associative theorizing, as expressed in unelaborated claims that "another related idea seems to be . . ." A third is to appreciate the arguability of all assertions about what behavior is or should be, and the richness of the assumptions underlying them.

From psychologists, philosophers might acquire a feeling for the extent to which speculations about behavior need to be disciplined by data, however feeble and imperfect. Extreme skepticism may enable one to dispatch any study. Yet the vigor with which such critiques are pursued may require some consideration of what beliefs about behavior will come in the stead of those empirical conclusions that have provoked the skeptics' wrath. In the absence of data, anyone's word may be advanced as a law of behavior. Rhetoric rather than statistics may carry the day.

At times, one may prefer the half-understood data of the psychologist, at times, the half-baked ideas of the philosopher. Such a balancing act is needed for each reader to derive the personally relevant insights that Cohen and those he criticizes have to offer.

Can children's irrationality be experimentally demonstrated?

Sam Glucksberg

Department of Psychology, Princeton University, Princeton, N.J. 08544

The assertion that all normal adults are fully competent to reason validly raises an interesting developmental question. If we cannot demonstrate incompetence in adults, then how might we do so in children? Cohen argues that any observation of less than optimal reasoning in adults cannot be taken as evidence of incompetence. What such observations can reflect is one or another performance factor. Among the performance factors that may affect reasoning are education and intelligence. In contrast, children reason differently from adults, and Cohen assumes that they are less competent than adults. How might children acquire adult competence if we exclude intelligence and education as developmental mechanisms?

Cohen alludes generally to experience and to maturation as mechanisms of developmental change. It is difficult to imagine that an abstract reasoning competence might develop solely as a function of neurological growth of maturation. [Cf. Chomsky: "Rules and Representations" *BBS* 3(1) 1980.] It is far more likely that such a competence develops through quite complex interactions among a child's experiences and his physiological maturation over a relatively prolonged developmental span. The usual outcome of this kind of developmental mechanism consists of marked differences in both the rate and the final level of development. It is, therefore, reasonable to assume that adults with different experiences will differ in at least some aspects of their reasoning abilities. Such differences must, according to Cohen, be attributed to performance factors, such as education or intelligence. But are not these just the factors that enable children to acquire adult competence in the first place? If so, then what are the grounds for rejecting the notion that adults may differ from one another in reasoning competence, but accepting that same notion when comparing adults with children?

Of course, the appeal to differences in education or intelligence to account for differential reasoning performances among adults is a neat way to preserve the assumption of universal and equal competence. But this seems to rob the notion of competence of any interesting psychological content. Consider this simple situation: Persons A, B, and C are given a test of reasoning. B and C do equally poorly, while A does quite well. A is a normal adult, educated in Western Europe, while B is also a normal adult who has not had any formal schooling, having grown up among a tribe in a South American rain forest. Cohen must assume that A and B are equally competent to reason validly, but that A has one or another performance advantage. C is a normal four year old, brought up either in the rain forest or in Western Europe. C is judged to be not as competent as either A or B. How does C become competent? He must simply grow up, either in the rain forest or in Western Europe. It shouldn't make any difference.

Perhaps it won't make a difference. There is something quite appealing to the argument that all people, irrespective of education or experience, will acquire reasoning competence simply because of their adulthood and humanity. In this respect, Cohen's competence assumption is analogous to Chomsky's (1968) assumption of universal and innate grammatical competence. However, there is a fundamental difference between grammatical competence, which is ascribed to all people, and Cohen's reasoning competence, which is ascribed only to adults. The grammatical competence possessed by the young child is quite abstract, and its primary function is to enable children to acquire language. Cohen denies such an abstract reasoning competence to young children, yet children must surely have some such competence if they are to use inference to acquire linguistic knowledge, and if they are to use inference to acquire the ability to reason validly in a wide variety of domains and contexts. It thus seems reasonable to extend the assumption of basic reasoning competence to children as well as to adults. As Cohen himself argues, no empirical evidence could, in principle, demonstrate incompetence in adults. This argument must be fully applicable to young children as well. We must then conclude that infant or child irrationality cannot be experimentally demonstrated.

If this seems unreasonable, then we might consider another alternative. It seems likely that many people do not use optimal reasoning strategies in a wide range of situations. Some people do use more optimal strategies than others in some domains. Do these differences among individuals, regardless of age, reflect differences in basic competence at some level of abstraction? We will not be able to answer this question until we can agree upon a coherent definition of "basic" competence. I suspect that when we have such a definition, young children will be shown to be competent. Yet they will not be as competent as the most competent adults. We should then be able to demonstrate relative levels of

irrationality between adults and children, and among adults as well.

Human reasoning: Can we judge before we understand?

Richard A. Griggs

Department of Psychology, University of Florida, Gainesville, Fla. 32611

It should be made clear from the outset that many researchers engaged in studying human reasoning do not interpret the behavior of their subjects as having "bleak implications for human rationality." In fact, it is my contention that the main concern of most reasoning researchers is not (and should not be) the question of whether they can or cannot demonstrate human rationality but rather the provision of a description and explanation of the observed behavior. I will return to this point, but first let me substantiate my first statement.

If the deductive reasoning literature over the past decade were examined, one would find that one of the most popular views of human reasoning (especially categorical reasoning) is represented by the rationalist school of thought led by Mary Henle. As Evans (1980) points out, this approach has been so influential that the two most recent collections of papers on human reasoning (Flamagne 1975; Revlin & Mayer 1978) "have been effectively dedicated to her cause" (p. 229). In Henle's forward to the Revlin and Mayer volume, she states (ironically in the context of some claims that I made about faulty reasoning behavior of subjects in a sorites task), "I have never found errors which could unambiguously be attributed to faulty reasoning" (p. xviii). These do not seem to be the words of a person despairing about the irrationality of humans. In fact, I have not found reasoning researchers to be worrying about human irrationality. Thus, the quotations with which Cohen begins his paper are really not reflective of the reasoning research area, at least as I perceive it.

Above I content that most reasoning researchers are not concerned with experimentally testing for rationality. If this is so, then how can I justify the *rationalist* school of thought? I do not see this as a problem. What has been labeled the rationalist school of thought argues that human reasoning behavior can be explained without resorting to "nonlogical" explanations, such as the atmosphere theory (e.g., Woodworth & Sells 1935). Thus, followers of this school feel that reasoning errors are brought about by such variables as the failure to accept the logical task, misinterpretation of premises, and the like. Part of the information-processing system postulated to account for the behavior employs rules that have been accepted as "rational." These rules, however, may be employed on a faulty data base. The prime example of such an explanation would be Revlin's model for categorical reasoning (e.g., Revlin & Leirer 1978). Even in this case I do not perceive the researchers as arguing that the experiments are testing human rationality but rather that the behavior may be brought about by an information-processing system that has a "rational" component. What about the misinterpretation of the premises, the failure to accept the logical task? Are these rational behaviors? I maintain that reasoning researchers are not really addressing such questions. Such judgments of rationality would seem, at least at the present time, not to be of primary importance.

It is difficult to judge a system as rational or irrational without first understanding the system. I do not perceive the progress of reasoning research to be advanced to a point that would permit such judgments. The main concern of reasoning researchers should be to gain an understanding of the cognitive functioning underlying the observed behavior in reasoning experiments. Formal systems, such as propositional logic and statistical theory, may be employed in the design of the experimental tasks, but subjects only make responses. The responses themselves are neither correct or incorrect. We can define them as such according to formal systems, but it is the reasoning researcher's job to explain why the responses were made and not to judge their rationality. The use of formal systems as components in such explanations (e.g., the rationalist school) does not entail a judgment of rationality on the researcher's part.

One recent criticism of cognitive research has been that it lacks ecological validity (e.g., Neisser 1976). Cohen seems to hint at this in his discussion of "cognitive illusions." For example, in the section of Wason's four-card selection problem, he states, "It seems, therefore, that experimenters' power to generate an illusion here depends on the relative unfamiliarity and artificiality and their apparatus." Cohen distinguishes between two types of realism; one using descriptive words or sentences on cards instead of letters and numerals, and the other using real objects (e.g., envelopes instead of cards). He then argues that for the first type of materials, subjects may still make the wrong choices in the selection problem (Manktelow & Evans 1979) but for the second type they hardly ever do (Johnson-Laird, Legrenzi & Sonino Legrenzi 1972). With real objects (the envelopes), the experimenters' power to generate an illusion supposedly disappears.

Such an explanation depends upon the replicability of the results that Cohen is using to support his argument; and, to be fair to Cohen, he clearly states at the outset of his discussion of the experimental work in reasoning that he is assuming that "in every case . . . the phenomena reported are replicable." In this particular case, however, they are not; and thus, Cohen's argument is not supported. We have twice attempted in our laboratory to replicate the Johnson-Laird et al. experiment and were completely unsuccessful both times. Merely using concrete objects like addressed, stamped envelopes instead of cards does not automatically bring about "logically" correct performance. Behavior in the selection task is more complex than that. This task is an excellent example of the struggle of reasoning researchers to understand their subjects' responses.

Evans (1980) has recently considered the question of ecological validity in deductive reasoning research. He argues that to understand real-life reasoning we must study it "ethologically, under real-life conditions" (p. 238). I agree that we should attempt to understand human behavior in laboratory settings and in more "real-life" environments, but in both cases we are trying to describe and understand behavior and not to judge it. Judgments of rationality could be made in either setting after we understand the information-processing mechanisms underlying the behavior. Reasoning behavior in laboratory conditions may be different from reasoning behavior in real-life conditions; but in both settings, humans are engaging in reasoning behavior. Both will provide useful information about human reasoning. Our goal should be to develop an understanding of such behavior in as many environments and for as many tasks as possible.

My final point was made recently by Newell (1980), but I would like to underscore it. Substantial literatures exist for deductive reasoning, probabilistic reasoning, decision making, and problem solving. Yet, in looking closely at any one of these literatures, it is difficult to detect the existence of the others. What we find are fragmented attempts to understand human cognitive functioning. Newell (1980) has argued for a unified theoretical endeavor in which all these various areas of cognition are viewed in a problem-solving framework. I support this notion and bring it up in this commentary because I feel that such an enterprise is crucial to our understanding of human cognition. Let us first understand as

best we can. Afterwards, we can worry about Cohen's question; but if we understand, Cohen's question becomes rather moot.

Another vote for rationality

Mary Henle

Graduate Faculty, New School for Social Research, New York, N.Y. 10011

We do not discuss what we take for granted. The present decade is not one in which to expect serious discussion of such conceptions as: faulty programming of reasoning, malfunctions of information processing, deductive programs, and the like. This situation is unfortunate because – if we did not assume that the process was computation – we would add "process" to Cohen's important variables of competence and performance. May I, nevertheless, suggest that the computer metaphor masks problems of *how* people reason.

My remarks will be addressed to the second part of Cohen's paper. Since my experience is mainly with deductive reasoning, I will largely confine my comments to this area. They will serve to support, from a different perspective, Cohen's demonstrations that research purporting to have "bleak implications for human rationality" has, indeed, no implications at all for rationality.

There is an increasing body of research (see, for example, Revlin & Mayer 1978) that suggests that errors in evaluation or selection of conclusions in deductive problems may be understood, first of all, by examining the premises as understood by subjects. Whether errors remain that cannot be so understood, I do not know. Indeed, the categorical syllogism, which I have mainly studied, seems almost to be a device for misleading the subject. The difficulties inherent in the copula have been much discussed. Recently Meyer (1980) has shown that when the copula is replaced by a noncopular verb, performance on syllogisms is much improved; and some of Revlis's work (1975) may be interpreted in the same way. I have suggested elsewhere other sources of misunderstanding of the premises of categorical syllogisms (Henle 1978), and I would guess that hypothetical syllogisms present difficulties of their own. On the logical particles, most invite difficulties of understanding. The misunderstandings of "some" are well known. Wertheimer (1959, pp. 256–58) has pointed out that "and" and "not" may have different meanings in different propositions; they may be "empty" or else have structural meaning. What differences these various meanings make for the reasoning process is worth investigation. Does "if" mean "if and only if" to a given subject? Again, judging from its use in the conversation of educated adults, my impression is that "every" deserves study.

Cohen's paper continues in new directions the close examination of instances in which performance errors do not argue for logical mistakes – a major task, I believe, for the investigation of logical reasoning.

The author's excellent discussion of the illicit conversion may also apply to other problems of the logical performance of untutored subjects. It is an example of how the subject's (here putative) understanding of the presented material determines the outcome of the reasoning. Given the implicit context of the utterance as presumably understood, what at first appears to be an illicit conversion is no longer illicit. (See Henle 1978 for a similar analysis of subjects' willingness to draw conclusions from two particular premises.)

Cohen points out that, to determine with certainty what probability a gambler has in mind in making his estimates, it would be necessary to question him in such a way as to rule out possibly incorrect estimates. The same "characteristic indeterminacy" applies to other kinds of reasoning. In order to explicate all the premises a reasoner employs in a deductive problem, it is necessary to question him so as to give him an opportunity to alter his argument. As Mill remarked long ago, an individual "has it almost always in his power to make his syllogism good by introducing a false premise; and hence it is scarcely ever possible decidedly to affirm that any argument involves a bad syllogism" (1874, p. 560). Rather than direct questioning, other devices, such as asking subjects to choose among Venn diagrams, will have to be found, if we are to understand their premises.

Cohen gives too much ground, it seems to me, to the proponents of human irrationality when he allows empirical knowledge to enter into the determination of the validity of inferences in everyday reasoning. The empirical truth of a proposition is one thing, its deducibility from premises is another. If a person evaluates a proposition on the basis of the former, he has renounced the deductive task. This does not mean that he is incapable of performing it; it does not even tell us whether he understands or remembers that that is what he was asked to do. The question is again one of *what* the subject is doing, not the *result* he obtains. It is unnecessary to water down deductions in everyday reasoning by introducing nonlogical considerations.

"It seems . . . that experimenters' power to generate an illusion here [Wason's four-card problem] depends on the relative unfamiliarity and artificiality of their apparatus," remarks Cohen. It seems to me that the role of both artificiality and unfamiliarity requires further study. It is true, as Cohen illustrates, that artificial and unfamiliar situations may encourage error. This is often the case in situations that the subject regards as absurd, although, as Cohen points out, he can solve formally similar problems without difficulty in everyday life. It is also true that syllogisms stated in "symbolic" terms are usually more difficult than those whose terms are "concrete." On the other hand, there may be cases in which the very concreteness and immediacy of the problem's materials distract one from its logical properties; I wonder, for example, whether this is why Wason's "Thog" problem is difficult (1977, p. 126). And the artificiality of the syllogism offers one great advantage to the solver: The premises from which a conclusion is to be drawn are presented neatly organized. In much real life problem solving, the relevant material must be recognized, sorted out from other material, and brought together before deduction can proceed. With regard to familiarity, it is not only a help in problem solving but may, on occasion, be an enemy of new solutions. The two roles of familiarity need to be sorted out.

It must be added that presenting a logical problem does not guarantee ratiocination by subjects: They may guess, evaluate the material truth of propositions, apply learned rules, or whatever.

Who shall be the arbiter of our intuitions?

Daniel Kahneman

Department of Psychology, University of British Columbia, Vancouver, B.C., Canada V6T 1W5

Cohen's paper addresses a significant issue: What is the relation between the psychological study of human reasoning and the normative study of inductive inference? The issue arises at this time because of the current surge of interest in the psychological study of biases and shortcomings of intuitive reasoning. This trend should be seen in historical perspective: It is a reaction against the fashion of using normative

models of optimal performance as theories of actual performance, a fashion that had taken hold after World War II under the influence of the sciences of information and decision, and as a reaction against behaviorism.

The use of normative models in a descriptive role had many beneficial consequences, and some unfortunate ones. It tended to stifle psychological investigation of processes and mechanisms, because people's responses were "explained" by the fact that they were correct; it also tended to give any refutation of a normative-descriptive theory the character of an attack on human rationality. We are trapped by the ambiguity of denials: The statement that "it is not the case that people are always rational" merely rejects an extreme thesis that would attribute rationality to every belief and act. This statement, however, is easily misunderstood as claiming that people are never rational. Cohen apparently misread the psychological literature in just this way, and was prompted to a superfluous defense of the human race against accusations of "deep-level irrationality" that had not in fact been made.

There are two major thrusts in Cohen's target article: an argument to the effect that the rationality of intuitions cannot be questioned because rationality is ultimately rooted in intuition, and an attempt to discredit psychological studies of errors of reasoning.

Cohen makes much of the standard argument that the authority of normative theories ultimately derives from their appeal to intuition. Indeed, one of the criteria for a norm of thought and action must be that reasonable people will want to obey it. It follows that reasonable people must be able to recognize a rational argument, and in this restricted sense that they must be rational. But it is improper to argue, as Cohen does, from this general belief in human rationality to a belief in the rationality of any notion for which a majority can be found. For example, one can think that Cohen's defense of the base-rate fallacy is irrational without being committed to a position concerning the irrationality of people in general, or of Cohen in particular.

An obvious difficulty for Cohen's position is that there are intuitions that are stubborn but make no sense. He chooses to dismiss some of those as cognitive illusions and to defend others as normative, but neither solution is very satisfactory. A deeper difficulty with which an intuitionist treatment of normative issues must deal is that the criterion of intuitive appeal, although *necessary* for normative power, cannot be *sufficient*, since people often find inconsistent intuitions appealing. It is all well and good to say that "an apparent conflict always demands resolution," but it is an article of faith that a resolution can always be found that will have as much intuitive appeal as the originally conflicting notions. At least in the domain of probability and inductive reasoning, it is a common occurrence to be convinced by a statement, equally convinced by a contradictory one, and unsatisfied by any proposed resolution. Indeed, the prevalence of such situations is more troubling for believers in human rationality than mere failures to apply accepted standards. I have a strong intuition that the strong intuitions of a rational person should be consistent, but my own intuitions fail this test.

Cohen has nothing of substance to offer on these difficult issues, beyond a vague message of faith, charity, and authority. Faith that all inconsistencies are apparent, none real; charity in finding an interpretation of any person's judgments that will eliminate all inconsistency; the authority of experts in competence to arbitrate remaining difficulties. Faith and charity are good things, but I find Cohen's faith in the consistency of intuitions puzzling, and his charity excessive and misplaced. Thus, it is clear from Cohen's article that he and I share similar intuitions about the gambler's fallacy and the base-rate problem, but we do not have the same attitudes toward those intuitions. I admit mine to be errors, which makes them so (unless Cohen issues a ruling to make such confessions inadmissible), but Cohen prefers to view the same intuitions as part of his reasoning competence. Unless the resolution of conflicting intuitions is itself intuitively appealing, the type of interpretive charity that Cohen advocates could easily become a reward for obstinacy.

How such intrapersonal and interpersonal conflicts of intuition are to be resolved is a problem for which I see no easy solution. But Cohen does have one, or rather a sort of administrative program for constructing solutions. He simply proposes that a theory of competence be read off "appropriate normative theories," which are apparently defined by their passing unspecified tests of intuitiveness. Because no hints are given of how this is to be done, the question of who will do it becomes relevant. Who will specify the tests, select the appropriate theories, and do the reading off? It is disconcerting to be told that this function will be performed by practitioners of a new domain of study, located in psychology but apparently reserved for competent philosophers.

In the first part of the paper, then, Cohen improperly extends an argument for the rationality of Homo sapiens to the rationality of almost every human belief; he effectively ignores the central issue of conflicting intuitions, and offers as a solution to the normative problem the suggestion that experts in matters of competence be trusted with it.

The second part of the paper is, I think, less good. Cohen distinguishes four categories in psychological research, and assures us that he can classify any psychological study in one of these categories. I believe him. In fact, he could have done just as well with only two categories, namely pardonable errors by subjects and unpardonable ones by psychologists. And it does not much matter which of his four categories is used, since the system is largely arbitrary. For example, almost any mistake that can be described as a cognitive illusion can also be viewed as a failure in a test of intelligence or education. The scheme is designed to guarantee that at least one of the categories will fit. Subjects who quickly admit that they have committed the gambler's fallacy by mistake will be classified as having suffered a cognitive illusion. If they stick to their guns, the experimenter will be accused of having applied an inappropriate model. Cohen's categories can be used as a handy kit of invective in encounters with psychologists. He also offers subjects in psychological studies a handy kit of defenses that they may use if accused of errors: temporary insanity, a difficult childhood, entrapment or judicial mistakes – one of them will surely work, and will the restore the presumption of rationality.

The argument that defends the base-rate fallacy is a caricature of the position that any error that attracts a sufficient number of votes is not an error at all. Here Cohen argues that a patient and a physician, looking at the same test data and sharing all their information, should be allowed to have different probabilities for a diagnosis. Furthermore, he agrees that the physician, who uses long-run frequencies, will in the long run make more correct diagnoses (and presumably make more correct decisions) than the ensemble of individual patients who ignore these frequencies. Would Cohen really insist on being treated according to his probabilities, rather than trust his physician's? This is irrational.

Improvements in human reasoning and an error in L. J. Cohen's

David H. Krantz

Bell Laboratories, Murray Hill, N.J. 07974

Cohen's treatment of rationality fails to recognize that adult humans continue to learn and to improve. He grants that

human children are inferior reasoners by adult norms; but the same point can be made ab out human adults, in terms of norms that they themselves subsequently may attain, and that others may have attained already.

New cultural inventions, such as probability theory, also improve reasoning. The analogy with human linguistic competence and performance is inapt: Natural languages do not improve, they only drift and borrow, but natural reasoning does improve. To deal with this, we must posit that the relevant "reflective equilibrium" is wide, not narrow, and it is only an approximate equilibrium at that; on a long time scale, it evolves.

I would agree that errors of human reasoning can often be classified as cognitive illusions or failures of education. In fact, most errors are *both* of these. The most important practical question is how to educate people (ourselves) so that we can avoid many illusions while not sacrificing much in our ability to cope with the rapid stream of inferences that we make in everyday life.

The preceding comments concern the main substance of the paper. I cannot, however, refrain from commenting on Cohen's egregious errors in his discussion of Bayes's rule (part II, section 4). I surely do not wish to defend Bayesian methods, in general, or to criticize propensity notions of probability, in general, But if 19,000 people suffer from disease A, for every 1,000 who suffer from disease B, and if a certain test result (call it T) occurs in ⅕ of those who suffer from A and in ⅘ of those who suffer from B, then the 19,000 who suffer from A will produce about 3,800 instances of T, while the 1,000 who suffer from B will produce only about 800 instances of T. Thus, out of every 4,600 instances of T, about 3,800 will come from individuals having disease A and only about 800 from those having B. If I were presented with that information, and had myself produced test result T, I would surely opt for the A treatment, rather than the B one. The population base rate of 19 to 1 is every bit as relevant to my case as the 1 to 4 and 4 to 1 population rates at which A's and B's produce T's. Neither population rate has anything to do with *my* test result T; neither has any *causal* linkage to my disease. I might just as well ignore one kind of population frequency information as the other. In fact, I would ignore neither, but would combine them, by the simple Bayesian arithmetic shown above, to obtain the posterior rate: *given* T, there are 38 A's to every 8 B's; and this arithmetical result makes me want very strongly to be treated as A, not as B.

I suspect most people would react in the same way: Knowing only that they had produced a T, and that the ratio of A's to B's among all people producing T's is 38 to 8, they would decide on the A treatment. There are two points here. First, even if we accepted Cohen's ideas about probability, his reasoning about the disease example would be wrong: Both probabilities involved are population frequencies, and by his analysis, neither the 19 to 1 base rate *nor* the 4 to 1 and 1 to 4 test-outcome rates would be relevant to the patient's treatment decision. Second, the example points up the absurdity of Cohen's use of propensities. It is hard to see, given Cohen's position, how any medical data at all could be used to help decide on treatment, from the patient's standpoint. So Cohen's position is wrong. Frequency information does provide arguments (though not necessarily overwhelmingly strong ones) for how to handle specific cases. These arguments should not just vanish when other arguments, not based on frequency, become available. If people do fail to use frequency arguments, then their reasoning should be criticized.

The question is not whether to adopt frequency or propensity notions of probability but how different arguments should be combined. That's a tough issue, as shown by Shafer (1976). But that thorny issue doesn't arise in Cohen's disease example, because both pieces of information are simple population frequencies, and the only tool needed to combine them is simple arithmetic.

Intuition, competence, and performance

Henry E. Kyburg, Jr.

Department of Philosophy, University of Rochester, Rochester, N.Y. 14627

1. Intuition is a slippery notion; the harder one tries to grasp it, the more difficult it is to hold onto. Cohen argues that criteria of rationality must be based on the intuitions of ordinary people, of lay adults. In some sense (which I return to below) this may be true. As his references to *untutored* intuition, *ordinary* people, *lay* adults, *normal* human beings, a *given* community, and *untrained* people reveal, however, Cohen means to advance a very strong thesis indeed. He wants to take the intuitions of ordinary people as the basis of "a coherent system of rules and principles by which those same people can, if they so choose, reason much more extensively and accurately than they would otherwise do." In doing this, they should be concerned only with a *narrow* reflective equilibrium – they should make the minimum modifications and qualifications in this set of intuitions required to render them mutually consistent. (According to whose standards of consistency? Cohen's intuition tells him that systems of belief ought to be "consistent" – although what that means is not spelled out in detail, and it would presumably depend on some systematic standard of deductive soundness. In any event, it presumably excludes the affirmation of both p and not-p. But Cohen need look no farther than his neighboring continent to find individuals whose intuitions on this score differ from his.)

Cohen also admits, though, that "however often faulted in performance" it must be correct to ascribe to normal human beings a cognitive competence that corresponds to the normative theory (part I, section 4). On the one hand, we are to discover the rational intuitions of people by looking at their performance; and on the other, we are not to suppose that failures of performance (how can we tell a failure?) impugn the normative theory of cognitive competence. This seems very much like eating your cake and having it.

Despite his constant emphasis, in the first part of the target article, on ordinary, untutored, lay intuitions, Cohen makes clear in the later section entitled "Tests of Intelligence or Education" that many apparent cognitive failures are simply due to lack of knowledge or intelligence on the part of those who are subject to them. This and the passage quoted earlier from part I, section 4 suggest that the intuitions of ordinary people can be guided by and even replaced by systems of rules and principles. If this is so, then we may properly ask for some guidance as to how one passes from the intuition embodied in actual (often faulty) performance to the competence into which people can be tutored.

2. One step consists in the systematization of the intuitions culled from performance. But systematization comes in many forms. Cohen refers to the criteria that are "the product of philosophical argument for some . . . wild reflective equilibrium" and contrasts this with "the narrow, bootstrapping, reflective equilibrium" which he thinks more appropriate for assessing the "rationality or irrationality of untrained people." But of course there are no "untrained" people; people differ in the degree and amount of training they have, and it is not at all clear that the rationality of people should be assessed by examining only people trained in public schools, and not people trained in universities. It is easy enough to understand the temptation to consider only narrow reflective equilibria, however (except that this does not solve the problem of deciding when they are "consistent"). It results from

the difficulty of finding a philosophical argument for a wide reflective equilibrium. That difficulty, in turn, reflects the fact that the last word has not been said concerning vast areas of rationality – particularly inductive and probabilistic rationality, but also rationality as applied to deontic, modal, or counterfactual argument.

Nevertheless, a lot has been said. We have systems of mathematics and geometry and set theory – not to mention systems of first order logic – for which good philosophical arguments *do* exist. Cohen points to the fact that even in sentential logic it is possible to find "obviously . . . invalid" arguments that can be given valid formalizations. But this is an argument for being careful about transcribing ordinary language into formal symbolism and vice versa. Of course, when people say "If p then q" they sometimes mean "$p \leftrightarrow q$"; logic instructors have always had to be careful to avoid such usages in constructing problems for their classes. Much ordinary argument, and practically all technical formal argument, can without strain be represented in the systematization provided by first order logic. There are also arguments that cannot be so represented, and it is an open question, often, how best to represent those arguments and how best to systematize their representations. This is just to say that deductive logic is not a finished system. When it comes to probabilistic reasoning, inductive reasoning, reasoning from empirical data to generalizations, hypotheses, and theories, the situation is even more open. There are a number – a relatively small number – of beginnings of systematizations vying for attention and approval in the market. There are also a reasonable number of general principles about which there is a fair degree of consensus.

3. Cohen gives the impression that what we need to do is to take a broad set of actual judgments and systematize them. But (to take the most successful case) this is not the way in which we arrive at the systematization of deductive argument provided by first order logic. On the contrary, progress in first order logic has consisted in the progressive *reduction* of the intuitions required for its basis. The model-theoretic proofs of the soundness and completeness of first order logic depend on a very few intuitions of a particularly compelling sort. It may well be that one can never dispense with intuitions altogether; but intuitions often can (and should) be dispensed with in favor of *arguments*.

It seems to me that the object of developing norms of rationality is not to honor all intuitions indiscriminately – even in qualified form – but to reduce collections of intuitions to a relatively small number of very basic intuitions from which others can be derived. The deductive intuition that q follows validly from p and if p then q seems to be such an intuition. This probabilistic intuition seems to be another: If r of the A's are B's, x is an A, and no reason can be given for thinking that x is more or less likely than any other A to be a B, then the probability that x will be a B is r; the appropriate degree of belief in "x is B" is r; the appropriate betting ratio is r:1-r; the appropriate number to put into the relevant computation of expectations or decision matrices is r; and so on.

4. What is the bearing of this on questions of the cognitive competence and performance of people? It seems to me that there are three quite separate questions involved. There is the question of the rationality of "ordinary people" in a variety of circumstances. It seems altogether natural (as Cohen agrees) that people should be more or less adept at cognitive performance. If, as Cohen suggests, it is fair of us to criticize children and animals for not measuring up to *our* standards of rationality, I do not see any reason why experts in rationality should not criticize ordinary people for not living up to *their* standards. We should hardly regard the system of mathematics as embodying inappropriate standards of mathematical rationality just on the ground that the vast majority of the people in the world can't add correctly. On the contrary, the point of developing standards of rationality – far removed though they may be from the ordinary performance of ordinary people – is, in some degree, to encourage an improvement in that performance.

There is also the question of what these standards of rationality are to be that are invoked in either the critical assessment or the constructive improvement of people's cognitive performances. The analysis of fragmentary accepted standards, the systematization of standards, and critical assessment of proposed standards, are all proper subjects for mathematical, logical, and philosophical inquiry. As Cohen shows forcefully, there are areas in which there is a lot of room for improvement of such standards of rationality as we have. This should not be taken as suggesting that in the course of history – particularly recent history – we have not made great strides in the articulation and refinement of systematic standards of rationality. Nor should it be taken as showing that the *vox populi* (of the educated middle class?) should be taken as the final arbiter of such standards.

It is a perfectly legitimate, and interesting, question to examine the extent to which people live up to the best standards of rationality we have; it is a perfectly proper social goal to attempt, insofar as it is feasible and worthwhile, to inculcate habits of thought and argument that conform to these standards; and it is a perfectly legitimate and important area of inquiry to attempt to develop yet more global and comprehensive systems of rationality, based on ever stronger and fewer intuitions of ever more limited scope. Indeed, the social goal may well be made more attainable by the very knowledge obtained from studies of the ways in which poeple fall short of current standards of rationality, as well as by the development of more coherent, comprehensive, and refined standards.

There is also a deeper question that Cohen may be addressing: the question of whether people are in some sense "intrinsically" rational, or have innate cognitive competence. I am not sure what to make of this question. Cohen seems to argue that no psychological tests or investigations can coherently reveal that ordinary people lack cognitive competence. Yet there is no doubt that such tests reveal that people often fail when it comes to cognitive performance. One would expect different communities (Cohen uses the word "community" only once, but he puts children and animals apart from the group with which he is concerned) to exhibit different sorts of defects. Perhaps he is suggesting that every human, qua human, can potentially achieve the same level of cognitive competence? The mention of Bernoulli and Euclid suggests that this isn't what he means, and if it were, it would still be a very strong and doubtful conjecture. It is quite obvious that standards of rationality, to the extent that we can develop them, are human standards, and that a standard that no human could approximate would clearly be inappropriate. But if we construe the investigation of rationality to be the investigation of the ways and the degrees to which various groups of people fall short of ideal rationality, we need not alter the ideal to conform to the performances of individuals – even of the most competent. We need not alter Peano's axioms because he couldn't balance his checkbook; even he may have had some bum intuitions about arithmetic.

Should Bayesians sometimes neglect base rates?

Isaac Levi

Philosophy Department, Columbia University, New York, N.Y. 10027

Tversky and Kahneman (1977) and others allege a widespread tendency among experimental subjects to make judg-

ments of probability that neglect base rates in violation of Bayesian norms of rationality. Cohen correctly notes that the neglect of base rates is sometimes legitimate. He appears to concede, however, that legitimate neglect may require deviation from Bayesian norms.

Cohen is right in suggesting that Bayesian norms are not always appropriate. But even if one endorses a Bayesian view of correct probabilistic reasoning for the situations that Kahneman, Tversky, et al. consider, the neglect of base rate information often remains legitimate. On this score, the experimental subjects reveal themselves to be better Bayesians than Kahneman and Tversky.

In his discussion of the task of assigning a subjective probability to a hypothesis as to the color of a taxicab involved in an accident, Cohen correctly distinguishes that task from the estimation of the long-run accuracy of the witness who reports the color in making such indentifications.

However, subjective probability assignments are often grounded in knowledge of objective, statistical probabilities in accordance with principles of direct inference. Thus, knowledge that the statistical probability of coin *a* landing heads up on a toss is p and that the coin is tossed on a given occasion might justify assigning a degree of subjective probability equal to p to the hypothesis that coin *a* will land heads up on that occasion.

Sometimes, however, the knowledge of statistical probability or long-run frequency needed for direct inference is not available, and other principles are invoked. The best known type of supplementary principle is insufficient reason, where equal probabilities are assigned to rival hypotheses when information favoring one alternative over another is lacking. It is well known that careless use of insufficient reason leads to inconsistency, and advocates of its use go to great lengths to avoid the contradictions – to no avail in my opinion.

Mistakes can also be made with direct inference. Kahneman and Tversky manage to commit some in their analysis of the taxicab example and kindred cases. The experimental subject can compute (using Bayes theorem) a long-run frequency of success of the witness's in identifying the color of cabs in the city along lines explained by Cohen.

But this chance cannot be used via direct inference to assign a subjective probability to the hypothesis that the witness has correctly identified the color of the cab as green. The reason is not that jurors are not concerned with the long run. That is true but irrelevant. The problem is that the jurors know that the cab under consideration was involved in an accident, and if they are to ground subjective probability on knowledge of long-run relative frequency, it must be on knowledge of the long-run relative frequency of the witness correctly identifying the color as green *of a cab involved in an accident*.

To obtain this information, the experimental subjects need to be told the percentage of blue (green) cabs in the city involved in accidents. This information is not given to them. It could be any value from 0% to 100%.

Hence, a good Bayesian should neglect the base rate given in the example because it is useless for the purpose of determining subjective probabilities.[1]

It is precisely in situations of this sort that principles of insufficient reason are invoked. If we assume that the experimental subjects reason like Bayesians, they proceed *as if* they supposed that 50% of cabs in the city involved in accidents are blue and 50% are green. But that is precisely what insufficient reason recommends.

Kahneman and Tversky have reported on variants of the taxicab experiment. According to one variant, the subjects are told that 85% of the taxicabs involved in accidents are blue and the remainder green. In that case, the subjects do not neglect base rates. That is not surprising; for in that case, the knowledge of statistical probability needed for direct inference becomes available. Insufficient reason is not invoked.

In another variant, the subjects are told that 85% of the taxicabs in the city (whether in accidents or not) are blue and the remainder green. The testimony of the witness is not reported. There, too, they rely on the base rate information.

Once more, this is not surprising. Direct inference is as illegitimate here as in the first situation. But insufficient reason seems to favor considering it just as likely that the cab involved in the accident is any one cab in the city as any other. This use of insufficient reason is inconsistent with the use of that principle in the first example. But we already know that persistent use of insufficient reason leads to inconsistency.

The remarks made here are grounded on strict obedience to Bayesian canons of correct probability judgment. They show that insofar as inconsistencies with these canons emerge, they appear to derive from a tendency to use insufficient reason when knowledge of statistical probability of the sort required for direct inference is unavailable. Since this practice is notoriously inconsistent, it is not surprising that it is revealed in the experiments under consideration.

But, as far as the issue of neglecting base rates is concerned, the experimental subjects studied by Kahneman and Tversky et al. seemed to have a better grasp of the matter – even from a Bayesian point of view – than do the experimental psychologists.

NOTE

1. The issue raised here concerns the misleadingly called "problem of the reference class." The topic is discussed in Levi (1977; 1980) where it is demonstrated that the general practice exemplified in the analysis of Tversky and Kahneman (1977) and preached by Kyburg (1974) violates fundamental canons of Bayesian reasoning.

Performing competently

Lola L. Lopes

Department of Psychology, University of Wisconsin, Madison, Wisc. 53706

Cohen's challenge to the current proclivity of some psychologists for reading "bleak implications for human rationality" into the behavior of naive subjects in laboratory tasks that require deductive or probabilistic reasoning is important and timely. I am in sympathy with many of the points that Cohen makes – although not with all of them – and I believe that the category structure he suggests for organizing the experimental literature on cognitive defects will be useful to those psychologists who are not already committed to the "bleak implications" viewpoint. But I have reservations concerning how useful it is for psychologists of thinking about human reasoning, in terms of the competence/performance distinction that Cohen proposes. In fact, it is possible that framing psychological analyses of human reasoning in such terms may exacerbate the already unfortunate tendencies of many psychologists to assume first, that only *poor* performance in reasoning tasks requires psychological explanation, and second, that the questions that are answered by normative theories are the same questions that arise in ordinary living. I will discuss these points in turn.

What needs to be explained? The entire literature on "human irrationality" can be taken to show that when subjects in reasoning tasks fail to perform in accord with some normative theory, psychologists feel obliged to explain why. But there is little corresponding interest in explaining those cases in which the behavior of naive subjects is exactly what it should be. For example, in the risky decision task in which subjects judge the worth of gambles (Anderson & Shanteau 1970; Shanteau 1974; Tversky 1967) and in the joint probability task in which subjects judge the likelihood of joint events

(Beach & Peterson 1966; Lopes 1976; Shuford 1959), subjects produce data that look "as if" they had been generated by multiplying. Since multiplying is the normatively correct response for these situations, the subjects must be doing something right. But *what* they are doing is not clear, particularly since most subjects do not know the multiplicative rules for expected utility and compound probability and since such subjects seldom report performing numerical computations of any sort.

Explaining how subjects multiply without multiplying is the kind of question that might be expected to interest psychologists. But I know of no research done on that topic save my own (Lopes 1976; Lopes & Ekberg 1980) in the more than twenty years since such multiplicative data first appeared in the literature. The problem, of course, is that when subjects perform correctly with respect to some normative theory, the theory itself seems to provide a ready explanation of their behavior: The subjects' data conform to the theory because their thought processes and symbol structures are somehow like the operators and variables that occur in the theory. Thus, subjects who produce data that look like expected utilities are assumed to be intuitive utility maximizers, and subjects who produce data that look like joint probabilities are assumed to be intuitive statisticians.

In developing the competence/performance distinction, Cohen seems equally prone to concentrate on performances that deviate – or seem to deviate – from a theoretical ideal. Little is said about the exercising of human competence save that it must be accounted for by the structure or design of human information-processing mechanisms. Yet it is this intuitive competence that demands psychological explanation, *not* as it exists in normative theories that have been systematized by logicians and mathematicians, but rather as it exists in the basic cognitive activities of ordinary people.

What are the real questions? In arguing for the competence/performance distinction, Cohen seems to accept the idea that normative theories necessarily answer the kinds of questions that are important to ordinary people. I am not so sure that this is true. For example, logic deals with the conditions under which patterns of argument are valid or invalid. People, however, are usually more concerned with whether conclusions are true or false. Thus, they often answer questions about the validity of arguments in terms of the correctness of the conclusion (Henle 1962).

I assume that Cohen would classify errors of this sort as cognitive illusions that most subjects would recognize as such on a few moments' prompted reflection. And I would not disagree. But I wonder whether emphasizing subjects' competence at recognizing the error of their ways might not obscure the functional significance that such "errors" have when subjects must evaluate unidealized arguments in the real world (involving, for example, vagueness, uncertainty, incompleteness, ignorance, and deceit) for which no normative theory holds.

When ordinary people reject the answers given by normative theories, they may do so out of ignorance and lack of expertise, or they may be signaling the fact that the normative theory is inadequate. How a psychologist interprets such behavior may depend on how he construes human reasoning abilities. For example, in behavioral decisions theory there has been growing sentiment over the last few years that current normative theories are unduly limited and unable to deal effectively with decisions made under realistic assumptions and constraints (cf. Einhorn & Hogarth 1980; March 1978). The fact that decision researchers have come to question the appropriateness of the theory rather than the expertise of the subjects probably reflects the fact that it has been many years since naive subjects were assumed to be intuitive statisticians or intuitive Bayesians.

It seems to me that it is healthy now and then to challenge the validity of normative theories – and Cohen seems to agree, given his warning concerning ascribing canonical authority to existing logic and statistics texts. But I wonder whether such challenges can reasonably be expected to occur if our assumptions concerning human competence for reasoning in accord with current normative theories keep us looking for the reasons that "performance errors" occur rather than asking whether something quite different might be going on.

Having concentrated in this commentary on explaining why the competence/performance distinction seems to me to have limited – and perhaps even negative – utility for psychological theorizing, I should not close without reaffirming the major usefulness that I see in Cohen's analysis of the categories into which research on human rationality can be sorted, particularly his sections on "misapplications of appropriate normative theory" and "applications of inappropriate normative theory." These sections show all too well that some of the bleakest implications to be found in the literature on human irrationality bear on the unseemly enthusiasm that some psychologists have shown for portraying human reasoning abilities in the worst possible light.

"Is" and "ought" in cognitive science

William G. Lycan

Department of Philosophy, Ohio State University, Columbus, Ohio 43210

I find myself largely in sympathy with Cohen's approach and have no serious quarrel with his scheme for classifying allegations of performance error. Instead of criticizing his paper directly, I shall offer a few positive suggestions as to the nature of epistemic warrant, and then briefly apply these to his argument against the possibility of a "probative wide reflective equilibrium" in inductive logic (part I, section 3).

It is an acknowledged but considerably underappreciated fact that the key notions in logic and epistemology are normative through and through: justification, warrant, legitimacy, license. To say that an inference is unjustified, unwarranted, illegitimate, illicit, impermissible, unreasonable, or irrational is to make a value judgment: It is to say precisely that the inference is one the subject *ought not* to have drawn. The "ought" here is not a moral "ought"; epistemic values appears to be sui generis, but it is *value* nonetheless. For this reason there is a prima facie difficulty in seeing how the property of epistemic warrant can be "naturalized" or located within the closed causal order that is the real world, exactly parallel to the much better publicized difficulty of locating the property of moral goodness within that same natural order. Epistemic warrant is no more obviously a matter of *fact* than is moral goodness.[1] And supposedly science treats just of fact, of what does happen in nature, and not of what ought to happen instead, or of what would happen in a better world than ours. This consideration lends real force to the concern expressed in Cohen's title. A psychologist who claims to have demonstrated experimentally that subjects exhibit systematic *irrationalities* is setting himself up in the always dubious business of deriving "ought" from "is."

Is there after all some naturalistically specifiable difference between what we call good reasoning and what we call bad or irrational reasoning, a difference that can be reconstructed within a genuinely empirical psychology? Armstrong (1973) and Goldman (1976 and elsewhere) have tried to bridge the face-value gap by explicating goodness of reasoning in terms of causal sufficiencies obtaining between belief states. By contrast, Cohen is trying to take some of the pressure off the naturalistic epistemologist by invoking a competence/performance distinction and assimilating theories of epistemic competence to theories of idealized physical entities not

actually found in nature. (See also Sober 1978.) Let me propose a third, alternative, hypothesis as to the natural ground of epistemic value.

There are certain deductive and inductive rules that we do in fact use in inference and that are (on pain of regress) fundamental rather than derived. Finding out exactly what rules these are and articulating them explicitly is an empirical task of Herculean difficulty, but we can be sure that even after it has been performed, we will still be able to ask of the fundamental rules, why we *ought* to obey those rules rather than others. Now, this question cannot properly be a request for a *proof* or derivation of the rules from any more fundamental principles, for by hypothesis the rules are themselves fundamental. The question must rather have the force of asking why it is good or desirable or useful for us to use those rules, to operate according to those principles, rather than others. And an illuminating response is available.

Let us suppose that we are the products of a benign developmental process that has our welfare, survival, and propagation at heart. (It does not matter whether we think of this process as the handiwork of God or simply as biological evolution. I will speak of Mother Nature.) And let us conceive of beliefs as being tools that we use in getting around the world and making life easier for ourselves. Now we may ask, what cognitive propensities would a benevolent Mother Nature have given us in order that we might be able to form the most useful beliefs? The first thing to grasp is that the sort of environment we occupy, our general shape and size, and the chemicals of which we are made jointly constrain Her further choices; She has only a very small space inside our skulls into which She must fit all the cognitive resources She can provide. She will be able to allow us only a small, finite stock of basic principles with which we may amplify and extrapolate from our immediate environmental input; these principles will have to promote great efficiency at some cost in care and detail. For example, She will build us to prefer simpler hypotheses to more complex ones, because they are easier to work with and afford plenitude of prediction out of parsimonious means (hence the canons of simplicity that implicitly govern straight-rule induction and explanatory inference; see also Sober 1981). Except in certain special cases, She will not sanction our changing our minds without reason, because the instability created by arbitrary changes in belief would be inefficient and confusing (hence the conservatism implicit in much of our epistemic practice). And so on. My hypothesis is, then, that our fundamental epistemic principles and habits, whatever ones they turn out to be exactly, are *good* principles, in that they are the ones that a wise and benevolent Mother Nature would have endowed us with, given Her antecedent choices of materials and overall anatomical structure. As Dennett (1978; forthcoming a; forthcoming b) observes, our having survived as long and in as handsome a style as we have proves that we are in fact well designed; our fundamental epistemic principles are as useful and beneficial as they can be, given environmental and anatomical constraints. I suggest that this is the "natural ground" of the normative notions of epistemology. "Rational inference," "best explanation," and so on ultimately reflect the optimality of certain human design features. We are fully rational when our optimally selected belief-forming mechanisms are working as they work when they are not flagging because they lack energy, are jammed by conative noise, or whatever. In this sense, I agree with Cohen that "we cannot attribute inferior rationality to those who are themselves among the canonical arbiters of rationality." [See also Pylyshyn: "Computational Models and Empirical Constraints" *BBS* 1(1) 1978.]

I have suggested that the rationality of an inference is at bottom the optimality of a design feature. But there are several ways in which Mother Nature's choices are at the same time suboptimal. First, if She had used better ma in the first place, She could have built a much more succe cognizer. We could have far greater storage capacity a better memories if we had bigger heads; if we were made o more durable hardware we would be subject to fewer malfunctions; if we had more sensitive receptors we would be less subject to perceptual illusion. Second, as Dennett (forthcoming a) remarks, Mother Nature is a *satisficer* and has made any number of design shortcuts. A human cognizer is a "passable jury rig," a "bag of tricks." No doubt some of our inductive rules are overly general in the interest of simplicity; a tendency to jump quickly to conclusions may be more useful on the whole than a respect for textbook adequacy of sample size; at the same time, as Dennett mentions, erring on the side of prudence in certain matters may be the best strategy. Third, proneness to false beliefs of certain sorts may serve important noncognitive evolutionary needs. We tend to overestimate the attractiveness and other admirable qualities of our own children; we quickly forget the painfulness of certain otherwise useful activities; we habitually deceive ourselves in any number of beneficial ways.

It is these three respects in which our cognitive design is suboptimal that make me more willing than Cohen is to grant the efficacy of a wide reflective equilibrium for inductive logic. He maintains that experiments conducted from the standpoint of *educated* induction or sophisticated statistical theory reveal only the extent to which subjects have or have not profited from training, rather than any built-in irrationality or erroneous competence. In one sense, I think, this is right: The fact that training and sophisticated reflection can improve our reasoning abilities and correct our preanalytical judgment does not impugn Mother Nature's craftsmanship – on the contrary. It does not show that She did not do the best job She could have done, given what She had left herself to work with. Thus, if my hypothesis as to the natural ground of epistemic value is correct, the fact of correctibility cannot reveal any built-in epistemic feloniousness. But in another sense it does reveal a *relative* cognitive inadequacy: We could have been better cognizers than we are if environmental conditions and raw materials had been different. With enough reflection, calculation, and discipline we may learn to recognize our specific cognitive shortcomings and to abstract from them when reasoning self-consciously. In this sense it is possible to achieve a wide reflective equilibrium that may differ nonnegligibly from Cohen's narrow bootstrapping equilibrium. Whether Cohen himself would agree with this I cannot say.

NOTE

1. It is interesting that this symmetry goes generally unperceived. Many many philosophers and laymen alike are scoffing moral subjectivists, relativists, emotivists, or nihilists, yet very few of these same people would think even for a moment of denying the objectivity of epistemic value.

Propensity, evidence, and diagnosis

J. L. Mackie

University College, Oxford OX1 4BH, England

In Cohen's argument about cabs and diseases (part II, section 4), everything turns upon the interplay between three kinds of "probabilities," namely frequencies, propensities, and epistemic probabilities, and the derivation of the last from the other two. "Propensity" may refer either to a nondeterministic causal tendency literally present in each single case (as the propensity of a radioactive atom to disintegrate may be), or to a statistical property resulting from a scatter of inputs to processes each of which is deterministic (as is the propensity

of a coin to fall heads; Mackie 1973, pp. 179–87; 1974, ch. 9). The propensity of a witness to give reports of various sorts about cab colours depends on the frequency of each colour of cab. This witness and the cab frequencies in this town together constitute a setup that has, as the experimenters calculated, a stronger propensity to produce incorrect identifications of cabs as green than correct ones of cabs as green. This in no way conflicts with the fact that the witness, having been confronted with a green cab, has an 80% propensity to report it as green. Hence, although, as Cohen says, the cab-colour ratios neither raise nor lower "the probability of a specific cab-colour identification being correct on the condition that it is an identification by the witness," they do help to determine the (epistemic) probability of a cab-colour identification being correct on the condition that it is an identification of the colour *as green* by the witness. And it is reasonable for a juror to be guided by the latter conditional probability, not the former. Why should he restrict the condition in relation to which he states the probability that is to guide his verdict, excluding the relevant and known fact that it was an identification as green? Although a change in the relative size of the green fleet would not affect the accuracy of the witness's vision, it would affect the credibility of his testimony when he reports a cab as green.

The logical point here was made long ago by Hume. Quoting the saying "I should not believe such a story were it told me by Cato," he argued that however good the witness, it is almost impossible for a miracle report to achieve credibility (Hume 1975, "Enquiries Concerning Human Understanding," section 10). Although it would not, in the circumstances envisaged, be a miracle for a cab involved in an accident to be green, this is sufficiently less likely than that it should be blue to need more than an 80% reliable witness to substantiate it.

The diseases example brings out the issue even more clearly. Although Cohen argues elsewhere (Cohen 1977b) that the standard probability calculus should be replaced by a nonstandard one for certain uses in relation to evidence, here he explicitly accepts Bayes's theorem; what he disputes is the choice of prior probabilities. Whereas the experimenters assumed prior probabilities of 95% that this patient has A and 5% that he has B, Cohen says that the prior probabilities are equal.

Cohen implicitly admits that he would use the unequal prior probabilities if we were told that among people who share all the patient's relevant characteristics 95% have A. But unless we are told that, he says, we have no reason to assume unequal predispositions to A and B in this patient: He refuses to assign unequal prior probabilities merely because there are unequal frequencies in the population. Such refusal would be legitimate if we had positive knowledge that this patient is in some relevant way not representative of the population. But if we have no specific information about him that points either to equal or to unequal predispositions, we are thrown back on the principle of indifference or insufficient reason, and the question is how we are to apply this principle. Cohen is (in effect) saying: "Focus on the individual patient; since there is nothing pointing either way about him by himself, we must regard the prior probabilities of his having A and of his having B as equal." The experimenters are saying: "Since we know that this patient is a member of the population, and don't know anything that relevantly differentiates him within it, we must take the prior probabilities as proportional to the frequencies of A and B in that population." And surely they are right. What is of practical importance, of course, is not the patient's abstract propensity to contract A or B, but his propensity to do so in the concrete situation, which will include whatever factors make A much more common than B in this population. Unfortunately this propensity is unknown. This patient *may* have a relative propensity to contract A as against B which differs from the 19:1 average relative propensity of members of this population. But neither he nor his doctor knows this, and the data stated give them no ground for assuming that his propensities for the two diseases, in the concrete situation, are equal. So in relation to what information they have, they must take the prior *epistemic* probability that he has A as 95%.

Cohen admits that an administrator with a long run of patients would regard the probability of A as greater than that of B even where the test indicates B, and would therefore dispense with the test. But this commits him to a paradox. Suppose that for none of the individual patients is there any positive evidence that he is relevantly nonrepresentative. Then Cohen is saying (i) that for each patient for whom the test indicates B it is more likely that he has B than that he has A, but also (ii) that for a large class of patients, for each of whom the test indicates B, it is very likely that nearly all of them – strictly, 19 out of 23, or 82.6% – have A. Each patient, he says, should demand the treatment for B, but the administrator should give them all the treatment for A. But how can this be, since the administrator is surely concerned only with the survival of all the patients, and there is no conflict of interest between one patient and another – for example, no competition for scarce resources? (The reference to "minimal cost" is a red herring, since Cohen would still say what he does if the cost of the test were negligible). Cohen's answer, therefore, cannot be defended within the Bayesian framework that he here allows.

Cohen suggests that the subjects, who tended to agree with his answer, were thinking rationally. It seems more likely that they failed to distinguish between these two statements about the test:

(i) When it is applied to people who have B, it is right 80% of the time.

(ii) When it says that someone has B, it is right 80% of the time.

If (ii) were true, it would be sensible to follow the indication of the test and opt for the treatment for B. But in fact only (i) is true; it does not entail (ii), and (ii) is false. Applied to members of this population with these symptoms, when it says they have B the test is right only 17.4% of the time. And all that each patient relevantly knows, however keenly interested he may be in himself as an individual, is that he is a member of a class (members of this population diagnosed by the test as having B) about which the test has only this degree of success. Since it is easy to confuse (i) with (ii), the tendency of the subjects to agree with Cohen's answer cannot be taken to express an authoritative intuition about the real force of the evidence in such cases as these.

The irrational, the unreasonable, and the wrong

Avishai Margalit and Maya Bar-Hillel

Department of Psychology, Hebrew University, Jerusalem, Israel

A normative theory of logical or probabilistic reasoning is acceptable, says L. J. Cohen, "only so far as it accords, at crucial points, with the evidence of untutored intuition." Such untutored intuitions, therefore, setting the standard that defines normativeness, can never be considered deviations from normative prescriptions. Ergo, if irrationality is taken to mean such deviation, experimental studies cannot demonstrate irrationality. Cohen allows, however, that experimental studies can demonstrate human fallibility, but insists that errors of performance do not reflect on our inherent competence for reasoning validly.

One could argue that many of the documented errors of

reasoning can be construed as deviations from normative prescriptions even by Cohen's own criterion. In other words, the deviations are from a theory with the proper credentials of agreeing *at crucial points* (which Cohen has left unspecified) with untutored intuition. But in any event, when errors in reasoning are both systematic and pervasive, it matters little to the investigator of human reasoning whether they are construed as deviations from an internally possessed competence, or from some external normative theory. Thus Cohen's thesis in itself leaves the body of research that he attacks intact. Far more disturbing is Cohen's attempt to justify some of the established errors. In some instances Cohen is not content merely to defend human rationality in the face of error, but would use these very errors as evidence of the superiority of intuition over supposedly normative prescriptions.

In what follows we question both the persuasiveness of Cohen's philosophical thesis, which we find an unreasonable one, and his reanalysis of one specific problem from the literature, which we find wrong.

Are there objective criteria for inferential validity? Let us consider first the case of deductive inference, which Cohen dwells on extensively. The justification of a rule of inference, we would like to claim, does not lie in its ability to meet some syntactic intuitions, but rather in its ability to preserve truth. Proper application of valid rules of inference guarantees that we never reach false conclusions from true premises. Cohen is not unfamiliar with this standard justification, but believes that "it does not come to grips with the whole of the underlying epistemological problem." He suggests an example of a formal argument which is valid in classical propositional calculus, but "which would obviously be invalid" once interpreted in natural language. His choice of example is most unfortunate. Far from finding it "easy to imagine situations in which [the argument's premise] is true and [the conclusion] is false," we found it quite impossible to imagine such a situation. To be sure, it is easy to imagine a situation in which

1. "If John's automobile is a Mini, John is poor and if John's automobile is a Rolls, John is rich"

sounds right while

2. "Either if John's automobile is a Mini, John is rich, or, if John's automobile is a Rolls, John is poor"

sounds wrong. But in that selfsame situation

3. "Either if John's automobile is a Mini, John is poor, or, if John's automobile is a Rolls, John is rich"

sounds right, while

4. "If John's automobile is a Mini, John is rich, and if John's automobile is a Rolls, John is poor"

sounds wrong. Indeed, it is hard to distinguish between the truth conditions of (1) and (3), or of (2) and (4). The mental exercise of imagining a situation in which (1) is true and (3) is false will either fail – in which case the intuition that (2) does not follow from (1) (which is false) will be indistinguishable from the intuition that (2) does not follow from (3) (which is true) – or it will succeed, in which case it will readily be seen that while (2) contrasts with (3), it does indeed follow from (1).[1]

Setting this particular example aside, it is well known that the natural language connectives "and," "or," "if then," do not always correspond exactly to their propositional calculus counterparts '$\wedge$', '$\vee$', '$\rightarrow$'. So obviously other examples can be generated, where untutored intuition concerning the natural language case would differ from the normative prescriptions concerning the corresponding propositional calculus case. Since the correspondence is only superficial, nothing is proved by such a deviation. But Cohen insists that the meaning of connectives cannot be acquired independently of our intuitions concerning the legitimacy of deductions in which the connectives are embedded. This is an unreasonably strong claim, however. Clearly connectives do not limit their appearances to inferences, and we see no reason why their understanding should hinge only, or even mainly, on their role in inferences.

Even if Cohen were to be granted his point, his account bears at most on how meaning is acquired and rules of inference are identified, but not on the justification of validity. Suffice it to imagine what would happen if some intuitively acceptable rule of inference failed to preserve truth: Clearly the intuition would yield to the truth-preservation criterion rather than vice versa. Cohen himself, when proposing his own example (part I, section 1) suggests that the inference there seems invalid because the premise seems true while the conclusion seems false.

Cohen aligns his thesis with one of Nelson Goodman's. But Goodman saw a two-way relation between intuitive inferences and normative inferences, with each affecting the other: "A rule is amended if it yields an inference we are unwilling to accept; an inference is rejected if it violates a rule we are unwilling to amend" (Goodman 1954, p. 67).

In probabilistic reasoning there is no analogue to the criterion of truth preservation. Nevertheless, here too there is a notion of consistency – to wit, compliance with the axioms of probability theory. The normative status of the demands for such consistency can be justified independently of intuitions by reference to the fact that only such consistency removes the threat of a dutch book being made against one. To be sure, this is a justification that is anchored in decision theory, where other justifications for other normative prescriptions for probabilistic reasoning are to be found. This commentary is not, however, the place to go further afield in discussing the justifications for probability theory as a model of rational reasoning under uncertainity.

Is it rational to have a policy of ignoring base rates? In spite of the consensus among many researchers that the errors that people are found to exhibit in judgment under uncertainty are dismaying (at the least), most if not all of these researchers would refrain from concluding that people are therefore irrational – though not necessarily for the same reasons put forth by Cohen. Thus, by directing his thesis against "the experimental study . . . of validity in deductive or probabilistic reasoning," Cohen is setting up a straw man.[2] We regard with greater alarm Cohen's inclination not only to excuse, but to actually condone, people's mistakes. For example, Cohen elevates the notorious gambler's fallacy into a "system" for beating the odds in some gambling situation!

We focus our rebuttal of this enterprise of Cohen's on his analysis of the base-rate fallacy in the cab problem, and the fatal disease problem, bringing both analytical and empirical objections to Cohen's treatment of these two problems.

Cohen reconstructs the Bayesian reasoning underlying the cab problem in the standard manner. However, while agreeing that "the ratio 17/29 is the value of the conditional probability that a cab-colour identification by the witness is incorrect, on the condition that it is an identification as green," he warns us that "jurors . . . or [other] people . . . ought not to rely on that probability." Instead, says Cohen, "if the jurors know that only 20% of the witness's statements about cab colours are false, they rightly estimate the probability at issue as 1/5."

In fact, however, the jurors know more than that 20% of the witness's statements about cab colors are false. They also know which particular color was mentioned by the witness on the particular occasion at issue. As Cohen himself shows, the proportion of times in which the witness's cab color identification is incorrect when it is an identification as green is expected to be 17/29 (59%), rather than 1/5 (20%). Moreover, while Cohen denies the relevance of this statistic concerning

long-run cab color identifications to determining "just the probability that the cab *actually involved* in the accident was blue" (part II, section 4, emphasis ours),[3] he does not hesitate to appeal to some statistic concerning a long run of witness statements to justify assigning 1/5 to the probability that the witness was wrong concerning the color of the cab *actually involved* in the accident.

It seems that Cohen considers 1/5 a more compelling estimate for the desired probability than 17/29 because the former, but not the latter, represents a causal propensity of some sort. But it is crucial to note that the causal chain is such that a green distal stimulus (i.e., cab) produces an 80% propensity of generating a "green" proximal stimulus (i.e., percept), *not* that a subjective sensation of green produces an 80% propensity of the perceived object being green. Formally, it is p (witness sees green | cab is green) which is the physiology-of-vision constant, while p (cab is green | witness sees green) depends on the ecological distribution of cab colors as well as on the witness's visual accuracy, as Bayes's theorem nicely shows (see also Bar-Hillel 1980). We wonder if Cohen would justify basing the probability that a cab is blue when the witness said it is green on the witness's probability of erring even in the following two extremes: (i) when the percentage of blue cabs in the city is 0% (i.e., when the cab color couldn't possibly be blue, no matter what the witness says); (ii) when the witness makes correct color identifications 50% of the time (i.e., when the witness does no better than chance, and so for all practical purposes is worthless).

Cohen addresses the role of base-rate considerations in the fatal disease problem as well. In this problem, an imperfect diagnostic test is to help in determining which of two possible diseases a patient suffers from. Cohen states his commitment to choosing a therapy on the basis of the test's outcome alone, rather than on the basis of the joint implication of the test outcome and the epidemiological statistics. The "standard statistical method" here, says Cohen, "is propagating an analysis that could increase the number of deaths from a rare disease." Unfortunately, Cohen's recommendation is propagating an analysis that could increase the number of deaths from the more common disease.[4]

We wish to emphasize that we agree with Cohen that in many instances causal tendencies are of more interest than mere concurrence statistics (see Jeffrey 1980). But in the two schematic examples Cohen analyzes, neither he nor the story itself provides any grounds for disassociating the propensity-type probabilities from the frequency type probabilities implied in the problems. Indeed, we urge the reader to consider Cohen's admission that "the administrator who wants to secure a high rate of diagnostic success for his hospital at minimal cost would be right to seek to maximise just that probability [presumably the probability of correct diagnoses over a long run of patients], and therefore to dispense altogether with the tests." Does the problem really justify that an individual patient's understandable concern for "success in his own particular case" should call for a medical policy different from the one dictated by concern for "stochastic success for the system"?

Cohen's defense of the base-rate fallacy encounters difficulties on empirical grounds as well. Consider a cab problem in which 85 to 15 represents not the ratio of blue to green cabs at large, but rather only of those in the records of local accidents. Surely if the first ratio "neither raises nor lowers the probability of a specific cab-colour identification being correct," then the second one shouldn't either. Hence if subjects were acting on Cohen's recommendation, they would ignore the base rate in this example as well. Such, however, is not the case. The distribution of responses to the latter problem was "radically different" (Tversky & Kahneman 1977, p. 176) from the one elicited by the original cab problem, and its median was substantially closer to the Bayesian answer.

Now Cohen cannot eat his cake and have it, too. He is faced with the necessity of acknowledging the existence of error in the pattern of responses to at least one of these problems (or in the rationale he provides for justifying the pattern elicited by the first).

The difference between the response patterns given to the two versions of the cab problem highlights another weakness of Cohen's account. According to that account, people's performance in such tasks is less than optimal either because they lack the competence to perform adequately (as in problems that only advanced education provides the necessary intuitions to deal with), or because "conditions are rarely, if ever, ideal for the exercise of such competence." How can such an account handle the fact that people with the same education and background, and operating under the same external circumstances, perform poorly on one task and adequately on another, nearly identical one (identical at least in the sense of making the same demands on the solver's competence)?

Conclusion. The position that rationality is defined by untutored intuitions is basically a metaphysical one. The observation that people, whether rational or irrational, whether tutored or untutored, often err when reasoning intuitively, is an experimental one. The experimental study of deductive or probabilistic reasoning is motivated by the desire to understand the cognitive processes involved in such reasoning. The implications of its findings to the image of man are extraneous to the study itself. The concern with erroneous intuitions has proven both convenient and fruitful, but is never an end in itself.

Cohen rises to the defense of intuitions that their propagators themselves are often willing to relinquish when confronted with their implications. Thus, for example, the gambler's fallacy, far from reflecting a "metaphysical belief" in "a spirit of distributive justice that regulates the whole cosmos," can be overcome by confronting people with the inconsistency between the fallacy and the belief in the absence of just such a spirit.

On the other hand, Cohen considers some intuitions too sophisticated to be held by "ordinary people, untutored in statistical theory," when in fact they do possess them, at least to some degree. Thus, for example, while the untutored clearly are not familiar with the law of large numbers in its mathematical form, they are aware that, ceteris paribus, a large sample provides a more accurate estimate of population parameters than a small one (Bar-Hillel 1979).

It seems that instead of debating the inherent rationality of humans, everyone stands to benefit from assisting them in overcoming some common and powerful biases in their intuitive thinking. So long as many decisions of monumental consequence to the individual and to society rely on the intuitive judgments of the unintelligent and the uneducated (which by Cohen's criteria most of us seem to be, see part II, section 2), we cannot rest on our laurels, content in the competence for rationality that Cohen guarantees us.

NOTES

1. Consider the statement "If John earns a million shekel a year, then John is rich, and if John lives on welfare, then John is poor." We handed subjects a page on which were listed versions (1) through (4) of this statement (analogous to (1) through (4) in the text). We asked them to indicate which of the statements was true and which was false, and whether any of them seemed to be saying the same thing. Of 17 subjects who marked (1) as true (6 others marked it false), 15 marked (2), (3), and (4) as false, true, and false, respectively. Eleven subjects stated explicitly that they thought (1) and (3), or (2) and (4), said essentially the same thing.

2. Cohen himself, in discussing the fatal disease problem, says: "It

is the former option that would be the irrational one for you, qua patient" (part II, section 4).

3. We gave subjects the following problem (after Lyon & Slovic 1976): "A certain factory manufactures light bulbs, 15% of which are known to be defective. To help in identifying the defective bulbs before they are marketed, the factory has a mechanical device that diagnoses the bulbs. This device identifies correctly 80% of the okay bulbs, as well as 80% of the defective bulbs. Consider 100 bulbs which the device has labeled 'defective.' What percentage of them would you think really are defective?" This problem is isomorphic to the cab problem, but asks for a percentage in a long run of correct identifications as defective, rather than for the probability that a particular identification as defective is correct. Thirty-one subjects out of a total of 52 gave 80% as their estimate.

4. Indeed, the following computation shows that Cohen's recommendation is likely to cost more lives. If the "standard statistical method" is employed, than all patients will be treated for disease A, the more common one, resulting in an average of 5% deaths (i.e., the 1 out of every 20 patients that actually suffers from B). If treatment is based on the test's outcome, then 20% of the patients will die, on the average, since the test gives the wrong diagnosis in 20% of the cases.

L. J. Cohen versus Bayesianism

Ilkka Niiniluoto

Department of Philosophy, University of Helsinki, 00170 Helsinki 17, Finland

Of the many interesting issues that L. J. Cohen discusses in his article, I shall concentrate on those raised in section 4 of part II. There Cohen argues that experimenters in the area of cognitive reasoning "risk imputing fallacies where none exist" if they test the subjects by applying an inappropriate normative theory of rationality. While this general thesis is no doubt correct, Cohen's defense of it is problematic.

Cohen's earlier criticism of the experiments of Tversky and Kahneman gives the background for his thesis (cf. Cohen 1979; 1980b; Tversky & Kahneman 1977). His main point is that the experimental results – which seem problematic for Bayesians – can be easily explained by assuming that the subjects apply what Cohen calls "Baconian probabilities" rather than the standard "Pascalian probabilities."

This issue is in fact closely related to an old controversy about the value of testimony on the "credibility of extraordinary stories" (Venn 1888, ch. 17). If p is the prior probability of a certain event, and if t is the probability that a certain witness tells the truth, then the posterior probability that the event did take place, given that the witness asserts that it has taken place, is

$$\frac{pt}{pt + (1 - p)(1 - t)} \qquad (1)$$

This formula, which is a direct application of Bayes's theorem, was known to Condorcet in 1785. One objection to formula (1) is the observation that in certain situations the testimony of a good witness will be "enormously depreciated": For example, if a reliable person (with t = .99) announces that the ticket number 267 has won in a fair lottery with 10,000 tickets, the credibility of this fact is, by (1), only 1/102 (cf. Todhunter 1949/1865, p. 400; Venn 1888, p. 415). Moreover, Venn suggests that "those who had made no study of Probability" would not treat this problem in the methodical way of (1): They would be at a loss in trying to choose between the opposite probabilities p and t (Venn 1888, pp. 418–19).

In the experiment of Tversky and Kahneman (1977), the prior probability p that a cab involved in an accident is blue is taken to be .15 and the reliability t of a witness who has identified the cab as blue is .80. According to formula (1), the probability that the cab in fact was blue, given the report by the witness, is

$$\frac{(.15) \cdot (.80)}{(.15) \cdot (.80) + (.85) \cdot (.20)} = \frac{12}{29} \approx .41 \qquad (2)$$

However, the most frequent estimate by the subjects in the experiment was .80, which is equal to the value of t. Tversky and Kahneman interpret this result as an error caused by the subjects ignoring the "relevant base-rate," that is, the information that the relative frequencies of green and blue cabs in the city are .85 and .15, respectively.

Cohen makes the important and correct point that we should distinguish here between the conditional probability that a cab-colour identification by the witness is correct, on the condition that it is an identification as blue, and the probability that the cab actually involved in the accident was blue, on the condition that the witness said it was blue. This difference is based upon a distinction between generic events (event types) and singular events; the former probability is relevant to a "long run of cab-colour identification problems" (just as Venn explicitly thinks in related situations), whereas the court should be concerned with the latter single-case probability.

While Cohen is thus right in criticising Tversky and Kahneman, his claim that the latter probability can be estimated "without any transgression of Bayes's law" is doubtful. A natural Bayesian interpretation of this situation is the following: The calculation (2) is wrong, because there are no compelling reasons to assume that the prior probability p is .15. In other words, there is no reason to suppose that the prior probability p (as a rational degree of belief) would be equal to the *relative frequency* of blue cabs in the city. This issue is related to the notorious problem of the choice of the right reference class: The relative frequencies of blue and green cabs might be very different in the whole country, in the local area of the city where the accident took place, and so on. Thus, it is not clear that the frequencies within the whole *city* have any special relevance for the problem. If this is so, then one strategy – which is traditional in the Bayes-Laplace school – is to divide the prior probability evenly between the two possible hypotheses on the colour of the cab. And if we choose p = ½ in formula (1), we obtain the result that the posterior probability is equal to t, precisely in agreement with the behaviour of Tversky and Kahneman's subjects.

This argument shows that the rationality of the probability appraisals of Tversky and Kahneman's subjects can be rescued by assuming that they are Bayesians who – in a quite old-fashioned way – employ uniformly distributed prior probabilities. It also follows that there is no need to impute "Baconian" rather than "Pascalian" probabilities to them.

Lay arbitration of rules of inference

Richard E. Nisbett

Institute for Social Research, University of Michigan, Ann Arbor, Mich. 48106

Cohen's article is very useful, I believe, for two reasons. First, the analogy between discovering normative rules of inference and discovering the grammar of natural language is apt in many ways. I would not carry the analogy as far as Cohen wishes to do, but I believe that working it through will prove to be an interesting enterprise with many benefits. Second, the article shows in high relief the nature of the problems that can result from making the untutored lay individual the arbiter of reasonableness in inference. Stich and I (Stich & Nisbett 1980) have argued that *any* such attempt is mistaken,

but I will not repeat those arguments here. Rather, I will argue that Cohen has not made a case for a lay standard of rationality at all (in fact, has not offered us any reasons why we should *want* untutored lay practice to provide the standard from which there is no appeal). Instead, Cohen has offered two contradictory lines of thought, one moving in the direction of no standards at all, and another implying that if education affects us in any way, it can only move us further away from rationality.

To illustrate the difficulties with Cohen's position, let us take a concrete problem of a type that most people would agree to be extremely common in everyday life (no parlor tricks) and see how Cohen might deal with the difficulties posed by people who happen to disagree with his solution: After the first two weeks of the major league baseball season, newspapers begin to print the top ten batting averages. Typically, the leading batter after two weeks has an average of about .450. Yet no batter in major league history has ever averaged .450 at the end of a season. Why do you think this is?

The answer that is preferred by most people who have been tutored in statistics (any by many who have not) is that two weeks is a relatively small sample of batting behavior and that deviant averages in such a small sample are much more likely than they are for the larger sample provided by the season as a whole. Suppose Cohen agrees with this answer. What can he do about a subject who persists in his assertion that "your so-called law of large numbers has nothing to do with it – it's just that the game favors batters over pitchers early in the season"?

Cohen acknowledges that there are such things as "cognitive illusions," so one might guess that he would simply maintain that this subject had fallen prey to such an error. But apparently this is not the position Cohen would wish to take since "nothing can count as an error of reasoning among our fellow adults unless even the author of the error would, under ideal conditions, agree that it is an error" (part I, section 3). By one reading, then, those lay subjects who refuse to admit their error are in fact not wrong. But of course, neither are the other lay subjects. Thus those who believe that the law of large numbers provides a satisfactory answer to the problem and those who do not are both right. Cohen is not likely to find many allies, lay or otherwise, if it is such a standard-free world he wishes to urge.

But perhaps Cohen would maintain that it is extremely unlikely that any subjects would continue to refuse to agree that they had made an error, "under ideal conditions." For most people, ideal conditions would involve education in the law of large numbers and, more broadly, in statistics and probability theory. But not for Cohen. On the contrary, such education in theory introduces a "bias," (part I, section 3), resulting in "substantially different conceptions of deducibility or probability from those once operating in their untutored judgments" and thus "the judgments of everyday reasoning must be evaluated in their own terms and by their own standards." [Cf. Warren: "Measurement of Sensory Intensity" *BBS* 4(2) 1981.]

So the ideal conditions necessary to get the subject to agree with Cohen do not include education. On the contrary, education is bias, and thus if the more educated person and the less educated person disagree, apparently Cohen would have to hold that it is the less educated person who is right. This at least avoids epistemic chaos, but at quite a price.

Perhaps Cohen means by ideal conditions just that the right answer is suggested to the subject and then the subject is allowed to reflect on his own answer and the proferred answer at leisure, so as to attain a new (narrow) reflective equilibrium. Cohen ought to be disturbed, then, by my empirical finding that just such conditions are sufficient to persuade most untutored lay people of the relevance of base rates for problems of the sort for which Cohen denies their relevance.

If "ideal conditions" refers to neither education nor reflection, then it would seem that Cohen's position has no clear content.

ACKNOWLEDGMENT

I am indebted to David Krantz, Stephen Stich, and Paul Thagard for comments on this reply.

Human rationality: Misleading linguistic analogies

Geoffrey Sampson

Institut Dalle Molle d'Etudes Sémantiques et Cognitives, CH-1205 Geneva, Switzerland

I accept much of Cohen's argument. He is surely right to say that it is circular to regard norms of reasoning as confirmable mathematically or empirically; their validity must be a matter of "intuition" in some sense. And I agree that some response patterns have been deemed irrational only because inappropriate norms were applied: I concede Cohen's part II, sections 3 and 4. But this leaves certain disagreements.

First, the psychological research discussed in part II is interesting because it reveals a paradox, that is, a pair of theses that both appear to be valid but that prima facie contradict one another – in this case, (i) the laws of thought derive their validity from human intellectual behaviour yet (ii) human intellectual behaviour often violates these laws. To reaffirm one side of a paradox, as Cohen does in part I, is not to resolve the paradox.

Second, Cohen's arguments rely heavily on analogies with Chomskian linguistics, and the aspects of linguistic methodology appealed to are questionable. This applies particularly to Chomsky's "competence/performance" distinction. It has often been observed (see, for example, Fodor & Garrett 1966, Moravcsik 1969) that this pair of terms is used equivocally by Chomsky and other linguists; by now it is perhaps no longer controversial to suggest that one of its chief functions is to protect linguists' theories from refutation, by permitting observations contradicting the theories to be dismissed as "performance effects," while those that confirm the theories are taken as a true reflection of a speaker's "competence." A clear case, discussed by Reich (1969 – and cf. Sampson 1975: 77–79), concerns the phenomenon of "self-embedding": Certain syntactic constructions that are normal enough when used singly produce bizarre, scarcely comprehensible utterances when iterated two or three times in a nested fashion, as in for example, "The boy that the cat that my wife found scratched was angry." It happens that Chomsky's grammatical theory predicts that such a sentence should be "good," since each individual construction it contains is normal, and the theory makes grammaticality depend exclusively on the properties of individual constructions rather than on the global patterns into which constructions are organized. Accordingly, Chomsky treats the unacceptability of such sentences as a matter of performance rather than competence. Yet it appears that the unacceptability of self-embedded sentences is systematic and is not predictable on the basis of any principles known independently of linguistic theory; the only basis for assigning the phenomenon to performance rather than competence is that this preserves the theory from refutation. (Reich rightly argues that we ought to prefer a different linguistic theory which predicts the "badness" of self-embedded constructions in a non-ad-hoc fashion, thus recognizing this phenomenon as a matter of competence.)

[See also Chomsky: "Rules and Representations" *BBS* 3(1) 1980.]

Cohen's discussion in part II, sections 1 and 2 might be thought to involve a circularity comparable to Chomsky's. What he says about, for example, the four-card experiments might be summarized as: Subjects intuitively observe *both* the law of contraposition *and* a rule making application of that law dependent on the familiarity of the materials involved in a given reasoning situation. On what grounds does Cohen treat the former as part of our competence but the latter as merely a performance limitation? – surely, only because to treat both rules as "competence rules" would damage his thesis that humans are not irrational.

If subjects in a reasoning experiment were presented with materials that were much less familiar even than cards bearing letters and numbers – if they perceived the experimental situation as some sort of science-fiction world in which all their commonsense assumptions broke down – then it could be that they would produce no answers, or randomly varying answers, to the experimenter's questions. In this case it might be appropriate to describe their performance as failing to reflect their reasoning competence because they were not in a position to apply their competence. (Similarly, my competence to distinguish grammatical from ungrammatical word strings in English will not be reflected in performance if the strings are written illegibly, say.) Or again subjects might produce no responses, or random responses, in situations in which the amount of computation needed to reach the correct answer was extremely large (just as my grammatical competence might not be reflected by my performance in response to enormously long sentences). But in the results discussed in part II, sections 1 and 2, subjects did not simply fail to reason and give random responses; rather, they systematically gave "bad" responses. It is not clear what (other than incompatibility with his theory) leads Cohen to dismiss these systematic responses as not being genuine reflections of subjects' reasoning competence.

The Piagetian point about logical competence not being expressly formulated by most people in a general fashion seems irrelevant, since it can hardly be supposed that ability to make a principle of rationality explicit is a necessary condition for being able to act in accordance with it. Similarly, in the case of language, it is not necessary and indeed not usual for a speaker who can use a given construction competently in his native language to be able to state the grammatical rules defining the construction. This problem may again stem from an unfortunate analogy with linguistics. Cohen uses the word "intuition" to stand for an unreflecting disposition to obey a given principle, but the word sounds as if it implies conscious reflection on the principle; Cohen expressly appeals to the precedent of "grammatical intuitions" in linguistics, but the term is quite equivocal in linguists' usage (Sampson 1979: Ch. 6).

Conditional probability, taxicabs, and martingales

Brian Skyrms

Department of Philosophy, University of California at Irvine, Irvine, Calif. 92717

Under "Applications of inappropriate normative theory" Cohen argues at length that in Kahneman and Tversky's taxicab problem, it is the psychologists rather than the subjects who have got the answer wrong. The mistake is held to have arisen from a confusion of two conceptions of probability: relative frequency and causal propensity. Since, as Cohen points out, an incorrect treatment of such problems could have serious practical consequences, it is perhaps worthwhile to look at the problem more closely, with special attention to the interaction of various types of probability.

Let us begin by introducing a third conception of probability into the picture; *credibility* or rational degree of belief, which is closely what is at issue for the mock jurors. Credibilities apply to single cases, via the singular statements that describe those cases. Information as to relative frequency or causal propensity has an effect on credibility, with the effect of propensity information being more dramatic. Given the value of a propensity, we will in general take that as the value of the credibility of the associated statement. Furthermore, we will take it as the value of the credibility even if we are given, along with the propensity, other frequency information. For example, the credibility of heads on toss 100 (given that the propensity of heads on toss 100 is ½ and that the relative frequency of heads in the first 99 tosses is 100%) is ½. The effect of relative frequency information is more modest. If we have no salient prior information about the outcomes in a finite class, our ignorance may be symmetric in such a way (exchangeability) that given the relative frequency of an outcome in that class and only that, we would take that value as the credibility of each statement attributing that outcome to a member of that class. For example, for the subjects of Kahneman and Tversky's experiment, it is plausible that the credibility of the statement that the taxi involved in the accident was green given that the relative frequency of green taxis among the taxis of Mudville is 85%, is 85%. There is a great deal more to be said about the interaction of these three kinds of probability (I have tried to say some of it in Skyrms 1980), but what I have said so far suffices for the analysis, and is, I trust, relatively uncontroversial.

Let us suppose that on the basis of the information that the relative frequencies of blue and green cabs are 85% and 15% respectively, the subjects adopt these numbers as their credibilities that the cab in question is blue and that it is green. Let us take the information as to the reliability of the witness as supplying the following information as to conditional *propensities:*

Propensity (the witness identifies the cab as green given it is green) = .8
Propensity (the witness identifies the cab as blue given it is green) = .2
Propensity (the witness identifies the cab as blue given it is blue) = .8
Propensity (the witness identifies the cab as green given it is blue) = .2

and accordingly, on that information the corresponding credibilities assume the same value. This information, in itself, would not plausibly change the prior credibilities as to the color of the cab involved. Finally, let us condition on the information that the witness identified the cab involved as green, to get the posterior credibility distribution. The posterior credibility that the cab in question is blue is, by Bayes's theorem, $(.85)(.2)/[(.85)(.2) + (.15)(.8) = 17/29$, not ⅕. All the probabilities involved in this application of Bayes's theorem are credibilities. Cohen's suggestion that a confusion of various kinds of probability renders the experimenters' reasoning invalid cannot be sustained.

Perhaps Cohen is interpreting the statement about the reliability of the witness as supplying information about the converse conditional causal propensities: The propensity that the cab is green given that the witness identifies it as green, and so on. Such is hardly a plausible interpretation, however. The color of the cab acts on the nervous system of the witness, not conversely. Suppose (to vary the example) that our prior information was that 100% of the cabs were blue. Would Cohen maintain that the cab in question had a .8 propensity to be green conditional on the witness identifying it as green?

I think that Cohen's discussion does establish the possibility that the subjects are involved in a probabilistic version of the fallacy of illicit conversion, moving from conditional probabilities going in one direction to those going in the opposite direction. This is not an implausible hypothesis, given that fallacies of illicit conversion occur in nonprobabilistic cases. I do not see, however, how conversational [implicature] can be brought in here to save the day for the common man.

Under "Misapplications of appropriate normative theory" Cohen offers a "pragmatic" defense for subjects thought to have been indulging in the "gambler's fallacy" or "fallacy of the maturity of chances." I find this discussion very puzzling. In it Cohen seems to be recommending a martingale strategy as a rational course of action. In the first place, I do not see how a predilection for martingale strategies is supposed to be mistaken by experimenters for a belief in the maturity of chances. In the second place, I am sure that Cohen is aware that for a gambler with a limited fortune, a martingale strategy constitutes as much of a gambler's fallacy as the fallacy originally at issue. (For a general treatment of gambling policies with stopping rules see Billingsley 1979.) As Cohen notes, in real life fortunes are always finite. What, then, is the point of his suggestion? Perhaps I have misunderstood the thrust of Cohen's brief remarks on this subject. If so, some amplification of them would be welcome.

Rationality is a necessary presupposition in psychology

Jan Smedslund

Institute of Psychology, University of Oslo, Blindern, Oslo 3, Norway

I regard Cohen's arguments and conclusions as essentially correct, but as suffering from a certain vagueness and incompleteness in some respects. I will try to reinforce his thesis by introducing some modifications.

Cohen sometimes appears to vacillate between a subjectivist and an objectivist position, that is, between a vocabulary referring to beliefs, judgments, intentions, and the like, and a vocabulary including expressions such as "malfunction of an information-processing mechanism." It should be clear that the question of rationality is meaningful only within a subjectivist framework – meanings and, therefore, implications and contradictions exist only for subjects. A causal mechanism as such is neither rational nor irrational. This means that when subjects fail on a task, the issue of rationality is meaningful only as long as the performance is described in terms of the subject's actual premises and conclusions.

Cohen relies heavily on the notion of intuition as a criterion of rationality. The question of how the validity of an intuition is established is answered by reference to consensus among adult laymen under optimal conditions of judgment. The optimality of conditions is, presumably, also a matter of intuitions. The suggested procedure defines the validity of an intuition in a given context by simple consensus. However, it does not establish the intuition as being necessarily, that is, noncontingently, true (in all possible worlds). I would, therefore, suggest that another criterion be added, namely, that the negation of the formulated intuition be absurd or meaningless. This would establish the intuition as true in all possible worlds. Although the negation test must also be based on consensus among normal adult speakers of the natural language involved, it may, in actual practice, be evaluated by the researcher with a few spot checks on available subjects. A particular subclass of such intuitions, based on proofs showing the negations to be meaningless or absurd, is described with the label "acceptable explanation" in one of my papers (Smedslund 1978).

So far, I have attempted to strengthen Cohen's argument by introducing the explicit presupposition of a subjectivist frame of reference and a check on the necessity of intuitions by negation. The next step is to reject Cohen's competence-performance distinction in favor of the simple assumption that all performance is rational. One difficulty of Cohen's distinction is that competence apparently must be described in terms of pure logico-mathematical operations, and that this leads to a reification of conceptual entities and an unfortunate ontological position that I have labeled conceptual realism (Smedslund 1977). There is not space to go into this here. However, the main reason the competence-performance distinction must be rejected is that it leads to the acknowledgment that performance is sometimes irrational. Cohen himself refers to Quine, acknowledging that "we have to impute a familiar logicality to others if we are to suppose that we understand what they say." However, he fails to acknowledge that this argument holds, not only for what others say, but also for what they do (Smedslund 1970). A consequence of this is that if the other individual's performance is not logical, then we cannot hope to understand him and are forced to resort to causal explanations in an objectivist framework. A simpler and more consistent position is simply to assume rational performance, that is, to assume that the other individual's performance is always a logical consequence of his momentary premises. This means that failures on experimental tasks demanding logical performance are interpreted as stemming from a lack of adequate communication between experimenter and subject. The most general reason for this is the experimenter's failure to take into account the subject's point of view, predispositions, previous learning, and limitations in capacity. As a consequence, the subject is acting on a set of premises different from those intended by the experimenter. In summary, it is suggested that, given a subjectivist conceptual framework, every act of every subject, including children and animals, must be interpreted as a logical consequence of the subject's momentary premises.

Having established the subjectivist frame of reference and the necessity of intuitions, and having abolished the competence-performance distinction, it is possible to classify all experiments attempting to demonstrate degree of rationality of subjects as *pseudoempirical* in the sense of involving attempts to test logically necessary statements by empirical methods. Cohen also (correctly) illustrates an important feature of such experiments, namely that they can be reinterpreted as testing the adequacy of procedures. An experimental procedure that results in apparent irrational behavior *must* be regarded as inadequate. In conclusion, I regard Cohen's work, with the suggested modifications, as a substantial contribution to a research program pursuing the question of what is necessarily true in psychology (Smedslund, in press a, in press b).

Some questions regarding the rationality of a demonstration of human rationality

Robert J. Sternberg

Department of Psychology, Yale University, New Haven, Conn. 06520

Cohen sets out "to show why recent research in the psychology of deductive and probabilistic reasoning does not have the 'bleak implications for human rationality,' as has sometimes been supposed." Even if this research does not have such bleak implications, there are aspects of Cohen's article that possibly do have bleak implications for human rationality. I will consider three such aspects.

Base rates. I found Cohen's arguments regarding base rates, and why in individual cases it often makes sense to ignore them, bizarre. Cohen seems to be arguing that the laws of probability work differently for individual cases than they

do for aggregates of cases. For example, he points out that "a patient is concerned with success in his own particular case, not with stochastic success for the system." He considers a case where "for a variety of demographic reasons disease A happens to be nineteen times as common as B." But he notes that "there is no information about you personally that establishes a greater predisposition in your case to disease A than to disease B. We have to suppose equal predispositions here, unless told that the probability of A is greater (or less) than that of B among people who share all your relevant characteristics, such as age, medical history, blood group, and so on." But if one were to pursue this line of argument to its conclusion, one would end up either with a sample size of one (oneself), or with so few cases that it would be difficult or impossible to estimate accurate prior probabilities in the first place. If one disease is 19 times as prevalent in the general population as another, should one really assume that the likelihoods are equal in one's own case, especially if, as in Cohen's example, the two diseases are equally dangerous in their potential consequences? The most reasonable strategy is to apply the population base rate to individual cases unless special circumstances, such as an unusual hereditary disposition toward a rare disease that shares symptoms with a more common disease, dictate otherwise.

Cohen's line of argument has been responsible for the maintenance of any number of essentially superstitious patterns of behavior. For example, classical psychodynamic therapy has probably survived as long as it has because of people's (and therapists'!) beliefs that despite the overwhelming statistical evidence against the method's general therapeutic effectiveness, the method will probably work in individual cases [See also Eysenck: "The Conditioning Model of Neurosis" *BBS* 2 (2) 1979.] "Miracle" cures for diseases such as cancer have survived for much the same reason: The desperate individual hopes that even though the treatment does not work on the average, it may work in one's individual case. After all, how could the "cure" have become popularized if it never worked? And of course smokers would have much greater difficulty continuing smoking, and soldiers facing enemy gunfire, if they did not believe themselves immune from the laws of probability that apply to everyone else. In some cases, such as that of the soldier facing battle, irrationality may actually serve a very rational purpose. But this shows only that irrational thinking can serve a rational purpose, not that the thinking is itself rational. Dreams, for example, serve a very rational purpose, yet are often irrational.

Deductive competence. Cohen claims that "in their familiar concrete concerns human beings show themselves well able to apply the law of contraposition to appropriate problems." This is a subset of a more general claim that people's reasoning processes may appear to be quite rational with concrete or everyday content, although not with abstract or artificial content. But does such a pattern of findings tell us something about people's logical competence, or about their ability to use knowledge of the world to supplement their logical competence? I would argue strongly for the latter claim. Consider, for example, the conditional, "If I drink a gallon of deadly poison, I will die." Most people would presumably recognize, correctly, that if I live, I didn't drink the poison. If presented with the same contraposition problem in terms of As, Bs, and Cs, the same people would, in most cases, fail to recognize the validity of the contrapositive. But this contrast seems to show that when the logical component is supplemented in its activities with real-world knowledge, people do not ignore the real-world knowledge, but can use it to their advantage. And it is not merely the concrete content of the terms that makes the difference, since nonsense syllogisms, with premises such as "If this object is a book, it is a typhoon," do not produce the same improvement in performance over abstract problems that factually accurate problems do. Cohen's own examples are generally open to this criticism. Cohen discusses the reasoning involved in chains of thinking such as "if the soap is not in the basin . . . it must be in the bath." But our reasoning processes are greatly facilitated by our past experience with soap and baths, and our knowledge about where we have kept soap in the past. This particular example, in fact, could be soluble solely on the basis of one's real-world knowledge without any recourse to reasoning. Abstract problems measure logical competence in the absence of a facilitative (or, in the case of counterfactual premises, possibly inhibitory) real-world component; and they show people's logical competence. Abstract problems do not show people's real-world "reasoning" abilities well, because real-world problems usually contain cues that facilitate the work of the strictly logical component.

The role of training. Cohen further claims that many of the fallacies people exhibit in their reasoning reflect nothing more than the absence of training in probability, statistics, or logic. This claim is no doubt correct in part. But the more interesting aspect of Tversky and Kahneman's work (1974) is that even extensive and intensive training doesn't seem to matter much: Mathematical psychologists attending a mathematical psychology meeting were prone to the same errors as laymen. Cohen claims that this only shows that "those who are supposed to have some statistical training . . . still fail to recognise new examples of regression to the mean for what they are." But the failure of these psychologists to recognize regression to the mean in new examples seems to tell us more about the psychology of thinking than about training. These are not just psychologists who have had "some statistical training." They include the world's experts in the field of mathematical psychology. If their training was not adequate, then whose would be? In fact, they supply the limiting case in which one *can* draw a conclusion about thinking rather than training, since they have more training in regression effects than practically anyone else, and also use regression concepts in their own research.

Conclusion. I have dealt with three of a number of instances in which I believe Cohen takes a basically sound point and attempts to push it much too far. His target article was potentially a good corrective: After the publication of striking findings such as those of Tversky and Kahneman (1974), Wason and Johnson-Laird (1972), there is an understandable tendency to go from one extreme to another – to assume that people are really quite irrational rather than really quite rational. Cohen correctly points out that it is easy to exaggerate people's irrationality. But, in certain places, Cohen goes to irrational extremes in his arguments, showing, perhaps, that people are not quite so rational as he claims.

Inferential competence: right you are, if you think you are

Stephen P. Stich

Department of Philosophy, University of Maryland, College Park, Md. 20782

The central move in Cohen's clever argument is the claim that the empirical theory of inferential competence must inevitably coincide with the normative theory of inference. The reason is that both theories exploit the same data, and in the same way. The data are intuitions about inferential validity, and in both normative and descriptive theories the goal is to construct the simplest and most powerful idealized theory that captures the bulk of the data. So the empirical investigator studying inferential competence *must* come up with the same account as the normative theorist. My disagreement with Cohen is easy to pinpoint. As I see it, he has simply

told the wrong story about the normative theory of inference. His narrow reflective equilibrium account of the way normative theories of inference are supported or justified cannot be maintained.

Perhaps the quickest way to underscore the problem with Cohen's account of the justification of normative theories of inference is to pursue the analogy, noted by Cohen, between grammatical theory and the empirical theory of inferential competence. Both a grammar and a theory of inferential competence are based on the data of intuition, and both are idealized to produce the simplest and most powerful theory that accounts for the bulk of the data. But as Cohen tells the story, there is one striking and paradoxical disanalogy. In grammar we expect different people to have significantly different underlying competences which manifest themselves in significantly different linguistic intuitions. The linguistic competence of a Frenchman differs radically from the linguistic competence of an Englishman, and both differ radically from the linguistic competence of an ancient Babylonian. Less radical but still significant are the differences among the linguistic competences of an Oxford don, a Shetland Island crofter, and a cockney fishmonger. Cohen, however, says remarkably little about the possibility or the extent of differences in inferential competence. At one point (part I, section 4) he suggests that *within* a community all normal persons would have the same inferential competence. His scattered remarks about "innate ability" and the abilities of "all normal adults" suggest that he thinks there is little or no intercultural difference in inferential competence either.

Cohen's reticence on the question of differences in cognitive competence is understandable enough. For the story he wants to tell about the grounding of normative theories would have considerable plausibility if there really were an innate inferential competence shared by all normal human adults. Yet there is substantial reason to think that inferential competence is far from uniform, either within a given culture or across cultures. Within our own culture, for example, surely education makes a difference. Subjects with training in logic, probability theory, and statistics are better intuitive reasoners than subjects lacking such training – not flawless, but better (see Nisbett & Ross 1980). If this were not so, logic, probability theory, and statistics would have no sensible place in a nontechnical education. Of course, Cohen might insist that education merely eliminates performance errors, allowing actual performance to approximate underlying competence more closely. But it is hard to see how he is in any position to insist on this a priori. The view has about as much to recommend it as the parallel suggestion that in teaching Eliza Dolittle to speak the English of the aristocracy, Henry Higgins was simply eliminating performance errors and enabling an underlying aristocratic linguistic competence to shine through. In the intercultural case, such data as we have are notoriously hard to interpret (Berry & Dasen 1974; Cole & Bruner 1971). Yet Cohen's own remarks suggest that we should expect intercultural differences in inferential competence. For he holds that "to ascribe a cognitive competence . . . within a given community is to characterise the content of a *culturally* or genetically inherited ability, which under ideal conditions, every member of the community would exercise" (part I, section 4, emphasis added). To the extent that the competence is culturally inherited we should expect people from different cultures to manifest different cognitive competences.

Now, if it is true that different people have different cognitive competences, then Cohen's account of the justification of a normative theory of reasoning faces considerable embarrassment. Recall that on his account of normative theory of reasoning is identical with a descriptive theory of cognitive competence. So if there are many cognitive competences abroad in our society and others, then there are many normative theories of cognition. But if there are many normative theories of cognition, which is the right one?

Though Cohen does not address this question head on, there are hints that he would endorse some version of individual or cultural relativism. In the case of apparently conflicting intuitions, he maintains, the people involved might come to recognize some tacit misunderstanding, or they might repudiate some previously robust intuition, "or they might conclude that different idiolects or conceptions of deducibility are at issue" (part I, section 1). Presumably the point of this last option is that if different conceptions of deducibility are at issue, then the problem of differing intuitions or competences is defused. For each person's (or culture's) intuitions may be right for him (or right for it). Right you are, if you think you are! But surely this relativist view is even more unpalatable for the normative theory of cognition than is its analogue in ethics. We are not in the least inclined to say that any old inference is normatively acceptable for a subject merely because it accords with the rules that constitute his cognitive competence. If the inference is stupid or irrational, and if it accords with the subject's cognitive competence, then his competence is stupid or irrational too, in this quarter at least. If he shares this irrationality with the majority of his compatriots, so much the worse.

By way of conclusion, let me note that there is a variation on Cohen's narrow reflective equilibrium story that does a much better job of making sense of our normative judgments about reasoning, both in everyday life and in the psychology laboratory. It seems clear that we do criticize the reasoning of others, and we are not in the least swayed by the fact that the principles underlying a subject's faulty reasoning are part of his – or most people's – cognitive competence. We are, however, swayed to find that the inference at hand is sanctioned or rejected by the cognitive competence of experts in the field of reasoning in question. Many well educated people find statistical inferences involving regression to the mean highly counterintuitive. But sensible people come to distrust their own intuitions on the matter when they learn that principles requiring regressive inference are sanctioned by the reflective equilibrium of experts in statistical reasoning. In an earlier paper, Nisbett and I tried to parlay this observation into a general account of what it is for a normative theory of reasoning to be justified (Stich & Nisbett 1980). On our view, when we judge someone's inference to be normatively appropriate or inappropriate, we are comparing it to (what we take to be) the applicable principles of inference sanctioned by expert reflective equilibrium. On this account, there is no puzzle or paradox implicit in the practice of psychologists who probe human irrationality. They are evaluating the inferential practice of their subjects by the sophisticated and evolving standard of expert competence. From this perspective, it is not at all surprising that lay practice has been found to be markedly defective in many areas.[1]

NOTE

1. Much of this commentary is based on the more detailed discussion in my "Could Man Be an Irrational Animal?" (to appear).

L. J. Cohen, again: On the evaluation of inductive intuitions

Amos Tversky

Department of Psychology, Stanford University, Stanford, Calif. 94305

The paper under discussion is the fourth publication in which L.J. Cohen attacks the work of Kahneman and myself. Since his position has shifted from paper to paper, it is instructive to

review his arguments briefly. Initially, Cohen (1977b; 1979) argued that people's tendency to predict by representativeness and to neglect (or underweight) sample size and prior probability should not be viewed as errors of judgment because, he claimed, such a tendency is consistent with his new calculus of Baconian probabilities, which has normative status equal to that of the standard theory of probability. Thus, we were accused of misclassifying some common patterns of judgment and inference as erroneous, relative to the standard theory, when they should have been classified as valid relative to Cohen's theory.

An examination of Cohen's theory, however, reveals that his normative claims are unfounded, his descriptive assertions are false, and his theory of inductive probability is therefore unacceptable from either normative or descriptive standpoints. For example, Cohen maintains that a conjunction is as probable as its less likely component. Let $P[A|X]$ be the (inductive) probability of A given evidence X, and consider a simple case where $P[A|X] = P[B|X] = \frac{1}{2}$ hence $P[A \text{ and } B|X] = \frac{1}{2}$. Now suppose that, on the available evidence, the probability that A is longer than B is ½ and the probability that B is longer than A is also ½. According to Cohen, then, the probability that A is *both* longer and shorter than B is also ½. Or, if the probability that the defendant killed the deceased is ½, and the probability that the defendant did not kill the deceased is ½, then the probability that the defendant *both* killed and did not kill the deceased is also ½. This is not a caricature; Cohen actually assigns positive inductive probability to self-contradictory propositions (Cohen 1977b, p. 222). On the other hand, if the probability of the conjunction of two events is set at zero, then, according to Cohen, only one of them can have a positive probability. Thus, if we assume that the defendant cannot be both guilty and innocent, and there is a positive probability, however small, that the defendant is guilty, then the inductive probability of innocence must be zero, contrary to both legal usage and common sense.

Cohen also insists that $P[A|X] > P[A]$, that is, an item of evidence X can increase but not decrease the inductive probability of a hypothesis (Cohen 1977b, p. 220). In this model, therefore, circumnavigation does not reduce the probability of the flat earth hypothesis, and an established alibi does not reduce the probability that the defendant is guilty. One can derive many similar anomalies in this system, but the point is clear. Cohen has produced a new concept of Baconian probability that differs drastically from the notions of probability that are used in everyday discourse, in courts of law, or in science. Cohen presents no valid arguments, descriptive or normative, to support his model. Instead, he claims that "this inductive form of probability fits the judicial discourse *perfectly*" (1977b, p. 45) and that "the experimental method of reasoning in modern science seems to have an essentially Baconian structure" (Cohen 1979, p. 393). And although intuitive judgments of probability in general and our experiments in particular contradict Cohen's model (e.g., evidence reduces the judged probability of an event as often as it increases it) he continues to claim that our experiments "merely confirm" the hypothesis that people reason in accordance with Baconian principles.

In our brief reply (Kahneman & Tversky 1979) to Cohen's earlier criticism, we exposed some of the defects of his system and invited the reader to examine Cohen's normative claims and to reflect on whether the tendency to neglect sample size or base rate should be viewed as mistakes that many of us are prone to make but would wish to correct or as opinions that should be maintained and defended. Cohen reacted by criticizing our appeal to reflective intuition. "We in any case expect, since we are no longer in the Middle Ages, that serious contributions to science should rest on deeper foundations than impressionistic appeals to intuition, common sense and ordinary usage" (Cohen 1980b, p. 89).

In the present paper, Cohen reverses his position on two basic issues, without warning or acknowledgment. First, after asserting that appeals to intuitions are worthless for the evaluation of the normative adequacy of a theory, he now argues at great length that intuitions are the only data by which normative issues can be settled. Second, after claiming that Baconian probability, rather than the standard calculus, captures the logic of controlled experiments and forensic proof he abandons the theory completely, again without an admission or an explanation. He introduces a new concept of propensity probability which, unlike the previous one, satisfies the classical probability axioms. Cohen does not tell us, however, whether the new system is intended to replace the Baconian system, and if so, why or whether it should supplement it, and if so, how.

Cohen's theories, we are told, are part of a factual psychological theory of inductive competence. Such a theory, I presume, should capture the patterns of judgments and inference that people find convincing or intuitively compelling. But how can such a set be characterized? People may endorse a given argument with confidence at first but reject it later, after several minutes, hours, or months of thought and reflection. Furthermore, human intuitions are often incomplete, inconsistent, and context dependent (see e.g., Kahneman & Tversky 1981, in press; Tversky & Kahneman 1981). The attempt to map and formalize human inductive intuitions, therefore, presents formidable empirical and conceptual difficulties.

Cohen proposes a simple solution. Using his own intuitions, rather than experimental data, he invents normative theories. If you are tempted to ask, How do we know that these intuitions are valid,? you will be told that "to ask for knowledge here is to ask for what is in principle impossible." The problem now is essentially solved because "accounts of human competence can be *read off* from the appropriate normative theories" (emphasis added). So empirical theories of competence are to be derived from armchair speculations. The division of labor between disciplines, according to Cohen, is clear. Psychologists should be confined to the study of (fallible) performance. Theories of cognitive competence will be "read off" by philosophers, "and though it is a contribution to the psychology of cognition, it is a by-product of the logical or philosophical analysis of norms rather than something that experimentally oriented psychologists need to devote effort to constructing. It is not only all the theory of competence that is needed in its area. It is also all that is possible." These quotes from the target article need no comment. Cohen's unquestionable normative intuitions are not only the basis for a cognitive theory of competence, they also set the limit for what such a theory could achieve!

The second part of Cohen's paper is a misguided critique of the psychological literature on judgment and inference. For example, he argues that people's failure to appreciate the effect of sample size on sampling variability is of little interest because naive subjects are not expected to know the law of large numbers. This argument misses a major point about psychological research. Of course, naive subjects are not expected to formulate or prove laws of statistics or geometry. However, the psychologist is very interested in whether naive subjects have learned from lifelong experience that nonrepresentative results are more frequent in small than in large samples. The question of which of the basic qualitative principles of statistics are represented in people's intuitions is not reducible to tests of intelligence or education.

Finally, Cohen misrepresents the psychological work, which he labels "the literature on cognitive irrationality." In fact, this work has been primarily concerned with the psychological processes that govern judgment and inference and has portrayed people as fallible, not irrational. Psychological studies show that human intuitions are sometimes wrong, and

that people often admit, upon reflection or discussion, that their failure to consider sample size or base rate was mistaken. In denying that inferences based on representativeness are prone to error, Cohen acts like a defense lawyer who tells his client "don't admit a thing, we will plead a Baconian interpretation of probability" even after the client has already admitted guilt.

The importance of cognitive illusions

Peter Wason

Psycholinguistics Research Unit, University College London, London NW1 2HE, England

The world is full of cognitive illusions. For instance, I sometimes suffer under them when I indulge in high level chess. It is not that I miscalculate, but that I misconceive the whole position. And I think that this sort of thing happens in a wider variety of spheres than is generally acknowledged, rather than merely being something that occurs at the behest of the sly psychologist. So I have no argument with Jonathan Cohen (other than to stress the generality of cognitive illusions) because, as he himself admits, none of us has claimed that my favourite one, the selection task or four card problem, in any way denigrates, or could denigrate, a competence theory of rationality. And yet I must correct one factual error concerning the alleged superficiality of the subjects' responses. In so doing I am not necessarily implicating the views of any of my collaborators; we have all developed rather different ideas about the relevance of the task.

There do seem to me to be some minor confusions in Cohen's account of our work, but I let them pass in order to avoid dreary technical niggles which would detract from my main point. Cohen asserts: "They [the experimenters] have manipulated the circumstances of a situation in such a way that subjects are induced to indulge in a form of reasoning that on a few moments' prompted reflection they would be willing to admit is invalid." But it is a cardinal feature of the four card problem that a fair proportion of subjects *conspicuously fail to correct their initial responses even when all the relevant information is made available to them to show that they are wrong.* This was first dramatically demonstrated by Wason (1969) and corroborated with improved controls by Wason and Johnson-Laird (1970) and by Wason and Golding (1974). The critical, unselected, and potentially falsifying card is frequently dismissed as irrelevant to the problem *after* it has actually been shown to falsify the test sentence, and admitted to do so. The subjects say things like, "It doesn't matter."

In one experiment (Wason & Johnson-Laird 1970) 74 percent of the undergraduate subjects (25 out of 34) failed to correct their solution after the falsifying card had been fully revealed and evaluated, and only 48 percent of these (12 out of 25) did so, even at a verbal level, after a twenty-minute clinical interview.

Now by any standards such behaviour seems irrational. The utterances of our subjects often resemble those of a hypnotised person. "The subject talks as if he were deluded; his attention is funnelled on his first decision; he contradicts himself in a way in which he would not do ordinarily; and he even denies the very facts which confront him" (Wason and Johnson-Laird 1972, p. 239). Our experiments have, in fact, been designed to probe the temporary effects of just such performance variables. And the most salient feature of performance is the subjects' incorrigible conviction that they are right when they are, in fact, wrong. (That is what the problem is all about, rather than being a pseudotest of how conditionals are understood.) When this conviction is contradicted by the facts, the subject is often presented with an acute conflict between two belief systems. This has analogies to real life crises of belief. Anyone who has campaigned for minority causes, such as unilateral nuclear disarmament, or witnessed the belief that love lasts for ever ("This is, and is not Cressid"), will know how ordinary people evade facts, become inconsistent, or systematically defend themselves against the threat of new information relevant to the issue. Cohen might retort that it hardly needs an experiment to demonstrate *that* kind of behaviour. But the point is to show that it can be manifested with innocuous material, and to explore ways in which the conflict might be resolved. Through experimental techniques we have the opportunity of simulating and ameliorating it.

On a miniature scale, such emotionally driven behaviour has been evinced from time to time in my experiments on reasoning since 1960, and it seems to me of much greater psychological interest than manifestations of rational competence outside the laboratory. The artificiality of experiments can be turned to good account by revealing factors that make people unreasonable: We do not uncritically extrapolate from them to real life, but (like an ablation experiment in physiology) we may learn from them something about the conditions that maintain rationality in everyday discourse.

There is something interesting, however, about the reaction of some people to my research. Jonathan Cohen is not the first academic to criticize it, but such criticisms have sometimes been rather affective in tone. In an earlier draft of the present paper Cohen referred to my experiments, not as cognitive illusions, which is splendid, but as conjuring tricks, which is a little derogatory. Others have been more impolite. Why? Those who are most concerned to vindicate the basic rationality of man seem to me a little worried by what might be construed as evidence to the contrary. Freud experienced this, and so, for that matter, did Picasso. Recently the four card problem was given (for reasons that escape me) to 801 French armed services recruits (Dumont 1980). Thirty-eight percent refused to attempt it, but their written comments on the test papers were enlightening; "C'est pourri, idiot, débile; c'est gaspi"* and "ça c'est belge une fois."** Perhaps some of my critics would joyously acclaim such sentiments as singularly apt. I think they support my view that the problem *is* often perceived as perverse because it seems to invite but defy understanding. The possibility of deception in obvious answers is unsettling. But it does not impugn rational competence.

EDITORIAL NOTE

* "This is rotten, idiotic, feeble-minded; it's a waste of time."

** "Now this is even Belgian" – the Belgians being benignly stereotyped by the French as getting everything backwards; for example, using the familiar form of address "tu" with strangers and the formal form "vous" with dogs.

Competence, performance, and ignorance

Robert W. Weisberg

Department of Psychology, Temple University, Philadelphia, Pa. 19122

Cohen's target article is valuable because it proposes an alternative interpretation for experimental results that purportedly demonstrate the pervasiveness of fallacies in human reasoning, and that have been cited as having important implications for the understanding of human reasoning. Given the acceptance of the work of Kahneman and Tversky (e.g., 1972b; 1973) and others by many psychologists (e.g., Glass, Holyoak & Santa 1979), Cohen's target article and past work (Cohen 1979; 1980b) have provided a valuable service in stimulating a closer look at the basic assumptions involved

in this work. While on the whole one may be sympathetic to Cohen's viewpoint, several questions are, however, raised by the present article. The first of these concerns the overall structure of Cohen's argument.

Cohen first attempts to demonstrate that rationality must be attributed to normal adult humans, because, to formalize a theory of logical reasoning in the first place, one must assume human rationality. Having established the "logical competence" of adult humans, Cohen goes on to analyze research results that have been interpreted as supporting the claim that human reasoning is defective in certain circumstances. Although I am not competent to judge the adequacy of the argument for the necessity of attributing logical competence to human adults, it seems that the data interpretations in the second part of Cohen's article are not at all dependent on the conclusions from the first part, as Cohen seems to believe. It is possible to disagree with Cohen's claim about logical competence and to agree with the main thrust of the second part of the article, that is, that human rationality may have been underestimated in various studies. From a psychologist's point of view, it is not clear why the first part of the article is necessary.

A second question concerns the competence–performance distinction which Cohen wishes to make. Cohen argues that the criteria that we use to evaluate human reasoning are founded on ordinary people's intuitions of deducibility. These intuitions are judgments that normal adults make in certain circumstances. If, therefore, an adult fails to make an appropriate judgment in some situation, there must be a performance problem of some sort involved. An example involves Wason's (1966) four-card problem. Subjects make errors in reasoning in this problem when the stimuli are cards with letters and numbers on them. When the stimuli are open or closed envelopes with stamps on them, fallacious reasoning is almost eliminated (Johnson-Laird, Legrenzi & Sonino Legrenzi 1972). Cohen explains these results by arguing that all subjects in these experiments possess deductive competence, but that competence is not generalized to the four-card problem, because that situation is too different from those earlier situations in which the relevant reasoning (use of the law of contraposition) has been carried out. Thus, the extension of one's logical competence to a new situation depends upon the similarity of content between that situation and previous situations.

The parallel that Cohen wishes to draw between logical competence and grammatical competence does not hold. The analysis of grammatical competence and performance is based on a distinction between the speaker's knowledge of rules of a language, and the psychological factors that play a role in the ultimate expression of this knowledge in the behavior of the speaker. Some of these psychological factors are memory limitations, changes in attention, and distractions. These, however, are not the sorts of factors that Cohen brings forth to explain why people commit fallacies in the four-card task. Cohen argues that a content difference is crucial in limiting the generalization of past experience to a new reasoning problem, which is not a psychological factor of the sort assumed to influence the expression of linguistic competence. Problems with Cohen's argument can be seen if we try to take his analysis of logical competence and formulate a comparable argument concerning language. Cohen seems to be saying that a speaker who has learned to talk about hockey will not be able to extend this competence to talk about the orbits of the planets, say, which is not the usual way in which performance factors influence the expression of linguistic competence. This introduces the further possibility that the postulation of logical competence may not be necessary for the understanding of logical reasoning. Indeed, if one accepts Cohen's content-based explanation for the failure to use the law of contraposition in the four-card task, then one has already accepted a noncompetence view, in the sense that Cohen is arguing against an abstract system of rules.

The final issue is of a different sort, and concerns questions of data interpretation. Cohen proposes four categories into which previous research results can be assigned without assuming that people commit fallacies of reasoning: cognitive illusions, tests of intelligence or education, misapplications of appropriate normative theory, and applications of inappropriate normative theory. The second category involves situations in which the alleged fallacy occurs because the situation involves principles of which nonspecialists would be ignorant. In such cases, as Cohen argues, judgments of fallacious reasoning are unwarranted. An examination of the examples placed in the other three categories suggests that several of them may more parsimoniously be interpreted as examples of ignorance on the part of the subjects.

One example of an illusion, according to Cohen, is Tversky and Kahneman's (1973) finding that words beginning with *re* are mistakenly judged to be more frequent than words ending with *re*. Tversky and Kahneman (1973) attribute subjects' errors to the heuristic of availability. Cohen argues that the term "heuristic" is misapplied here, because subjects are not using some sort of discovery plan, as "heuristic" implies. On the basis of his earlier discussion, Cohen argues that subjects must be granted the competence to avoid fallacies in reasoning about probabilities or frequencies, which includes some procedure requiring that the person check whether the available evidence constitutes a fair sample. The differential ease of recall of the two sorts of words in the judgment task circumvents the evaluation of the representativeness of the evidence.

An alternative explanation is simply to assume that the adult believes that the evidence is representative; one therefore does not have to make any claims about why procedures are not carried out. One need only assume that people who are ignorant of memory research do not know about the role of various sorts of bigram cues in determining the relative ease of recall of words.

As an example of misapplication of appropriate normative theory, Cohen discusses the alleged fallacy of illicit conversion, in which a person draws an inference from "if p then q" to "if q then p." This tendency has been attributed to people's unwarranted treatment of conditionals as if they were statements that allowed one to infer q from p. Cohen argues that one does not have to assume that people are making unwarranted assumptions, but rather that rules of conversation (Cohen's interpretation of Grice 1975) make it likely that an isolated statement of the form "if p then q" will be treated in such a way that conversion is permissible. In this situation it again seems possible that the prevalence of illicit conversion is due to ignorance on the subject's part. In the context of formal logical reasoning, a conditional statement is not interpreted in the same way as it is in the context of ordinary conversation. Explicit instruction is required for one to learn to compute the truth value for a conditional statement, given the truth values of its components. One would not therefore expect naive college students to reason correctly from such statements.

As an example of the final category, applications of inappropriate normative theory, Cohen discusses why subjects fail to take prior probabilities into account when judging the probability of a given outcome. He argues that the tendency to impute fallacious reasoning to subjects depends on the application of an incorrect model of probability as the basis for determining correctness of judgment. As an alternative, Cohen proposes a "Baconian" analysis of probability, which, if applied to such cases, would not result in the attribution of fallacious reasoning to someone who ignored prior probabilities. As in his other examples, however, it can again be argued that simple ignorance is involved here. If most subjects are ignorant of Bayes's theorem, then one would not expect prior

probabilities to have any systematic effect on judgments. Cohen discusses this possibility in passing, but does not pursue it, perhaps because he considers the presentation of his interpretation of probability to be of higher priority. However, discussions of alternative models of probability (e.g., Cohen 1970; 1980; Kahneman & Tversky 1979) may become superfluous if one concentrates more on what subjects know in these various situations.

In conclusion, in the target article and in his earlier work, Cohen has raised potentially important questions concerning the interpretation of a sizable body of experimental results. However, his interest in philosophical questions concerning foundations of logical theory and models for probability may have resulted in some over interpretation of the data.

Cohen on contraposition

N.E. Wetherick

Department of Psychology, King's College, University of Aberdeen, Old Aberdeen AB9 2UB, Scotland

My comments will be concerned mainly with Cohen's interpretation of the results obtained by Wason and his associates. Many papers have been published using the four card task, and I accept Cohen's view that they have identified what might be called a "cognitive illusion" analogous to "perceptual illusions." The perceptual illusions have also been the subject of much experimental work, but while it may be held that they are phenomena of interest in their own right, it seems to me that, by an extension of Cohen's arguments, Wason's illusion may be seen to be of relatively little interest.

Cohen's identification of the difficulty as "a failure to apply the law of contraposition" is almost certainly correct. Suppose the rule were "all ravens are black." The contrapositive equivalent of this rule is "all nonblack things are nonravens." So, logically, an observation of a nonblack nonraven confirms the rule in the same sense as does an observation of a black raven (Hempel 1945; see in particular p. 11 et seq.). This is of course counterintuitive; although there may be both an infinite number of ravens and an infinite number of nonblack things, most people feel that, in the world as it is, there must also be infinitely more of the latter than of the former, so that the contrapositive observation, although logically equivalent, does not (to say the least) confirm the rule as strongly as the direct observation. It might be argued that the reason that either the direct or the contrapositive observation confirms the rule is that the observation might have been negative. The raven might have been (but was not) nonblack, and the nonblack thing might have been (but was not) a raven. The same consideration applies when we consider an observation that was negative. Ravens that are nonblack are logically equivalent to nonblack things that are ravens, but we feel intuitively that there are infinitely more nonblack things than there are ravens, so that it must be at least uneconomic to set out to test a rule about ravens by observing nonblack things. Wason, however, requires his subjects to act as if it were as reasonable to observe nonblack things (i.e., not-q's) as to observe ravens (i.e., p's), in order to be judged correct.

It was always difficult to see why realism (in Cohen's first sense) should affect the issue, but Manktelow and Evans (1979) now seem to have shown conclusively that it does not. While I accept Cohen's argument that realism in his second sense abolishes the fallacy, Johnson-Laird, Legrenzi, and Sonino Legrenzi (1972) do not succeed in demonstrating the fact. They used fourpenny stamps and fivepenny stamps, which were at the time the minimum postage rates for unsealed and sealed envelopes. There are good grounds for believing that their subjects tested the real life rule "if a letter is sealed it has a fivepenny stamp on it *and if unsealed it has a fourpenny stamp*." Thus a sealed letter (p) was examined (it needs a fivepenny stamp), but an unsealed letter (not-p) was not examined (either stamp will do); a fivepenny stamp (q) will do for either a sealed or an unsealed letter, but a fourpenny stamp (not-q) is only sufficient if the letter is unsealed. This is a valid process of reasoning, leading to the choice of p and not-q only, but it is not the process that was supposed to be under investigation.

Wason (1960) is open to analogous criticisms. There the rule was "any three numbers are correct so long as they are in ascending order." This rule entails many less general rules, such as "three successive even numbers," "a number, twice the number, three times the number," and any set of numbers satisfying one of these less general rules also satisfies the correct rule. The contrapositive suggests only that numbers not in ascending order are incorrect – which is true but unhelpful. It requires considerable subtlety to question a rule that always generates positive instances, particularly in view of the fact that a prejudice in favour of rules of minimum generality is widely regarded as evidence of commendable scientific caution. Nevertheless, that is what Wason's subjects were required to do.

In three papers (Wetherick 1970; 1971; 1973), the last a review of Wason and Johnson-Laird (1972) I advanced criticisms of this work similar to those of Cohen, but they were ignored. Perhaps coming from a logician of Cohen's eminence more notice will now be taken of them. Both the four card problem and the series problem used in Wason (1960) deceive intelligent subjects into giving what is logically the wrong answer without telling us anything about the behaviour of such subjects outside the laboratory. There is, however, no doubt that the authors believed at first that they were contributing to a scientific understanding of intellectual function in real life. Later, as so often happens in psychology, the experimental task came to be regarded as an object of interest in its own right, and some recent contributors to the literature are scarcely aware why it was invented in the first place. This illustrates what is perhaps the most important general point arising out of Cohen's target article. Granted that formal logic (whether Aristoltelian or modern) and mathematics are the products of human intellectual function near its highest pitch of excellence, why should it be deemed appropriate to measure ordinary (intelligent but untutored) human beings against the standards set by logic and mathematics in order to condemn them as irrational when they fail? It is not difficult to design experimental tasks in which subjects behave in accord with the prescriptions of logic, but this of course is not news, and the experiments would probably not be published. It is a mark of the almost total failure of creative imagination among experimental psychologists that they so rarely attempt to come to grips with human intellectual function as it is, but clutch at anything that has the appearance of constituting an external objective standard. Such investigations serve no useful purpose because, as Cohen has pointed out, standards of rationality are and can only be set by human reason.

Unphilosophical probability

Sandy L. Zabell

Department of Mathematics, Northwestern University, Evanston, Ill. 60201

Cohen believes that normative theories must accord "at crucial points, with the evidence of *un*tutored intuition" (emphasis added), a view he also ascribes to Nelson Goodman. In doing so he seriously misreads Goodman. For what Goodman (1979, p. 64) really says is:

> Inferences are justified by their conformity to valid general rules, and . . . general rules are justified by their conformity to valid inferences. But this circle is a virtuous one. The point is that rules and particular inferences alike are justified by being brought into agreement with each other. *A rule is amended if it yields an inference we are unwilling to accept; an inference is rejected if it violates a rule we are unwilling to amend.* The process of justification is the delicate one of making mutual adjustments between rules and accepted inferences; and in the agreement achieved lies the only justification needed for either.

Our *un*tutored intuitions are contradictory, and any consistent normative theory must contradict some of them. In this light let us examine the base-rate examples discussed by Cohen.

1. Consider the example of the two diseases, A and B, that Cohen refers to. To make matters even clearer, let us suppose that (i) A is not merely 19 times as prevalent as B, so that $P(B) = 1/20$, but that $P(B) = 10^{-6}$, and that (ii) the frequency of a correct diagnosis is 51%. Suppose the test reports disease B. According to Cohen's logic, it is more reasonable to believe that an event of frequency 0.51×10^{-6} has occurred (the patient has disease B) than an event of frequency $0.49 \times (1 - 10^{-6}) \approx .5$.

Cohen contradicts himself here, since he first charges that "the literature under criticism is propagating an analysis that could increase the number of deaths from a rare disease of this kind," but then almost immediately concedes that a hospital following the procedure of always diagnosing A would maximize its rate of diagnostic success. And therein lies the absurdity of Cohen's position. He believes that an individual patient is better off concluding he has disease B, given the test so indicates, even though, under the frequencies postulated above, only about one in a million patients will survive such a choice. (If the frequencies postulated are less extreme, as are Cohen's, the same difficulty remains, albeit it is less striking.)

Now in fact, before the test, the probability that the test will give a correct diagnosis is indeed 51%; Cohen is confusing this prior probability with the posterior probability that a particular diagnosis is correct. This is somewhat akin to Fisher's famous fiducial error: The frequency with which a random interval covers a fixed but unknown parameter is *not* the same as the posterior conditional probability that a given interval contains the parameter (see, e.g., Good 1971, pp. 138–40).

2. Cohen argues that a "causal propensity analysis" indicates that we should take $P(A) = P(B) = 1/2$, even though the frequency of B in the overall population is known to be $1/20$ (or 10^{-6}), unless we know the probability of B for people who share all "relevant characteristics." This is an unwarranted application of the principle of insufficient reason, given that we do have information as to the population frequency. But moreover, people do not act as Cohen says they should; his explanation is inconsistent with Kahneman and Tversky's own results.

Take the cab example. Cohen says we should adopt equal predispositions "unless we are told that because of faulty maintenance, say, the probability of a blue cab's being involved in accidents that share all the relevant characteristics of the present one, such as poor braking, worn tires and the like is greater (or less) than that of a green cab's being involved." But Kahneman and Tversky in fact report that as soon as test subjects are told that "85% of cab accidents in the city involve Green cabs" (Tversky & Kahneman 1980, p. 63) – *all* accidents, *not* just those that share all "relevant characteristics" – the test subjects do take base rates (at least approximately) into account.

3. Cohen states that if Kahneman and Tversky's conclusions about such examples stand, then it would be "difficult to defend the continued use of lay juries." This is a worry Cohen has expressed before (1977b, p. 263), and he has also argued that his alternative interpretation of their results is preferable because "it does not imply such widespread and evolutionarily valueless stupidity among lawyers and other non-psychologists" (1977b, p. 262).

It is easy to find, however, instances in which common practice is often seriously in error, even among professionals; or the dubious portacaval shunt (Grace, Muench & Chalmers *et al.*, 1966), for example, Blondlot's nonexistent N-rays (Klotz 1980), or the embarrassingly high number of scientists who have been conned by Uri Geller (Randi 1975). And juries in particular are known to be highly imperfect instruments.

The parallel with the use of eyewitness identification is instructive. Although recent research has made it clear just how flawed this type of evidence can be (Clifford & Bull 1978; Loftus 1979), Münsterburg's pioneering work (1908) on this subject was already available at the beginning of the century. Yet the legal system still relies on eyewitness identification, for much the same reasons that it still relies on the jury system: It is – or at least was until an era of Nixon tapes and Abscam videos – the best one can do in most situations. The work of Clifford, Bull, and Loftus is valuable precisely because it alerts us to the limitations of eyewitness testimony; one hopes it will enable us to adjust for them.

And just so the results of Kahneman and Tversky: Their insights into how people go astray in making intuitive probability assessments have already found application, for example, in statistics texts (see Freedman, Pisani & Purves 1978, pp. A-12, 15, 24).

In his *Treatise of Human Nature* (1739), David Hume described certain types of "unphilosophical probability," and gave a primitive psychological description of how they could arise as a result of a conflict between judgment and imagination. We are indebted to psychologists such as Kahneman and Tversky for their efforts to replace Hume's tentative insights with those of a sounder, more scientific basis.

Author's Response

Are there any a priori constraints on the study of rationality?

L. Jonathan Cohen

The Queen's College, University of Oxford, Oxford OX1 4AW, England

I. The overall perspective. Those who write about experimental research on rationality have an understandable tendency to suppose that no controversial issues are raised by their choice or application of normative criteria for rationality. Instead they tend to concentrate on new or old experimental data and on the extent of the data's conformity to these criteria. Such writers may well be provoked, like **Weisberg,** to query the relevance of the philosophical remarks in part I of my target article to the interpretations of experimental data in part II, or even to suggest in general, as **Fischhoff** does, that philosophers and psychologists cannot usefully co-operate with one another in any way at the present stage. What I was trying to show, however, was that the actual interpretation of experimental data is bound to be affected by the resolution of certain fundamental issues about the

normative criteria for rationality. If the fundamental credentials of those criteria are as I claimed them to be in part I, then it follows, first, that any replicable defect in human performance must be put down to some situationally explicable malfunction or underdevelopment of mainline computational mechanisms rather than to the routine functioning of sideline "heuristic" mechanisms, and, second, that there are several important ways in which performances that seem at first sight to be defective may turn out not to be so.

It was no part of my purpose, as **Griggs** seems to suppose, to contend that *all* contemporary psychologists are "despairing about the irrationality of humans." On the contrary, I pointed out that diametrically opposed opinions are still to be found on this subject, and I suggested that at one extreme there is too much emphasis on competence and at the other on performance. Nor did I anywhere even accuse *some* psychologists, as **Kahneman** curiously misreads me, of claiming that people are never rational. But I did suggest, as do Nisbett and Ross (1980, p. 74), that an unflattering characterisation of the layman's capacities has become rather commoner in the past decade than in previous ones. And this is quite compatible with accepting the value of much previous work, such as that referred to by **Henle,** as constituting excellent vindications of lay rationality in the performance of certain deductive tasks.

Moreover, not only are psychologists, evaluating replicable laboratory data, still deeply divided about the extent of lay rationality, as the commentaries on my present paper suffice to show; so too are philosophers, evaluating one-off utterances outside the laboratory. Finocchiaro (1981) has argued rather cogently that many of his fellow philosophers are much too prone to impute fallacies to nonacademic people. Alternative interpretations can be easily constructed, he shows, for many of the texts that philosophers quote as examples of fallacies actually committed by journalists, politicians, advertisers, administrators, and the like.

Nor was it my purpose to theorise about rationality in general, so as to cover rationality in action, as is supposed both by **Mynatt, Tweney & Doherty*** and by **Einhorn & Hogarth.** "Rationality" is unfortunately rather an imprecise term, so in the first sentence of my target article I defined the sense in which I was concerned with it as "validity in deductive or probabilistic reasoning." And in this area it certainly seems rather paradoxical to claim, as do **Daniels & Smith,*** that correct performances "generally reflect the exercise of collective or social and not just individual critical faculties." Unless this is just a truism about the general guidance afforded by education, it is true only if the correctly performing adult normally submits every deductive or probabilistic issue to the deliberation and decision of his family, workmates, and so on. But that is manifestly false even in tribal or collectivist countries, let alone in western democracies: surely we very often estimate for ourselves – from our general knowledge and immediate perceptions – when it is most probably safe to cross a road, wade a river, climb a tree, or whatever.

II. The epistemological credentials of criteria for assessing the validity of laymen's deductive inferences or the accuracy of their probability judgments. I argued that the validity of criteria for lay inference need to be broadly in accordance with lay people's own intuitions. The objections to my argument fall into nine categories.

1. **Tversky** claims that the present argument is inconsistent with the criticism in Cohen (1980d) of Kahneman and Tversky's (1979) appeal to their own intuitions and those of their readers. However, I emphasised in the target article that the intuitions that can support normative criteria are "of those who are not theorists themselves." The trouble with Kahneman and Tversky's (1979) mode of reasoning was that it was intrinsically circular. Tversky is therefore rather obviously misrepresenting my position when he says that "using his own intuitions" Cohen "invents normative theories." On the contrary, I specifically repudiated such a procedure, and cited, "the practice of the courts," for example, as a source of evidence for the theory about probability in forensic reasoning that was developed in Cohen (1977b).

2. **Kahneman** objects that I am treating mere accordance with lay intuitions as a *sufficient* foundation for normative theory, and this objection is also implicit in the subtitle of **Stich's** commentary: "Inferential competence: Right you are, if you think you are." However, what I actually argued was something quite different. Common sense and ordinary usage are certainly not a sufficient foundation. First, good normative theories, like good scientific ones, should exhibit the derivability of a very wide variety of relevant hypotheses from a rather narrow set of fundamental ones. Particular hypotheses come to corroborate one another by their derivability within such a system (Cohen 1970, p. 86). The system building carried out by logicians and probability theorists is integral to the establishment of a solid normative theory. Hence if certain intuitions do not fit an existing model, that is a reason for either discarding the intuitions or constructing a new model; it cannot be a reason for giving up the search for a systematic model (and systematisation in these areas is bound to involve logical or mathematical formalisation). Second, a systematic logic of deducibility aims to be inferentially sound, in the sense that the conclusion of any inference that it licenses should be true if the premises are. This requirement is implicit in the first-mentioned objective of the system, since each of the actual or possible intuitions whose content it aims to systematise is about a deduction that, allegedly, has a true conclusion if its premises are true. We have here an indispensable minimum for what is meant by "deducibility." Similarly, a normative system for probabilistic reasoning must operate with a concept of probability that satisfies one fundamental constraint: it must be able to function as a gradation of inferential soundness (Cohen 1977b, pp. 13 ff.). One way of ensuring this is to establish an appropriate interpretation for the classical calculus of chance, as was achieved by Ramsey (1931) and the other authors I cited. Another way of doing this is to establish an appropriate interpretation for a suitably generalised modal logic, as in Cohen

(1977b).

3. The task of constructing adequate normative theories is therefore by no means an easy one; and, as **Kyburg** says, even deductive logic is not a finished system. Many issues remain notoriously controversial because the relevant intuitions are hard to reconcile. But the intuitions of which I speak here are in every case "intuitions in concrete individual cases," as I described them. It is against these that hypotheses about normative laws are to be tested, just as hypotheses about natural-scientific laws are tested against individual observations, and hypotheses about ethical principles are tested against the deliverances of conscience in real or imaginary quandaries. So **Mynatt, Tweney, & Doherty*** are wrong to compare my view with that of Thomas Reid (1969) and to object that I am invoking "'intuitions' of criteria for fallacies." (**Shafer*** makes a similar misinterpretation.) Some philosophers do indeed proceed in this way. For example, **Kyburg's** commentary appeals to "a relatively small number of very basic intuitions from which others can be derived." But such a procedure is inherently question begging. Theorists are out to establish general principles, and so their ultimate evidential data must be particulars (see Cohen 1982).

4. All this is perfectly compatible with different people sometimes having inconsistent intuitions about the same inference. In such a case the appropriate analysis may either be that marginally different conceptions of deducibility are involved, as in the Lewis and Langford (1959) versus Anderson and Belnap (1974) controversy to which I referred. Or it may be that someone has made a mistake. **Zabell** is therefore no doubt right to insist that not all untutored intuitions are mutually consistent. But he is wrong to think that this is an objection against what I wrote, or that I am misreading Goodman (1954). Normative theories do indeed require as much confirmation as they can get from people's concrete intuitions. But their normative role requires them to operate also as standards that reject conflicting intuitions as false.

5. **Smedslund,** on the other hand, objects that no one should ever be interpreted as having made a mistake in reasoning. According to Smedslund, if we are to suppose that we understand what others say, we have to impute logicality to them, not just as a normal practice but as gracing each and every one of their inferences. But this argument does not succeed. It must surely be allowed that people sometimes intentionally contradict themselves on a matter of fact, since they often admit to having changed their minds; and it is then difficult to deny that such contradictions may also take place unintentionally because of forgetfulness. So a pair of mutually inconsistent assertions by the same person may sometimes be quite well understood, on the assumption that each is separately self-consistent. And the conventions of interpretation that allow such mutually independent understanding for molecular units of discourse are all that are needed to allow the intelligibility of invalid inferences.

6. **Sampson** has a different kind of objection, though his point may also lie behind **Stich**'s and **Kahneman**'s objections. He argues that I offer no resolution for the paradox implicit in attempts to combine the view that "the laws of thought derive their validity from human intellectual behaviour," as he puts it, with the view that "human intellectual behaviour often violates these laws." But I offered no views at all about the actual source of validity for the laws of logic, and I certainly would not wish to identify this source with human behaviour. What I actually said was that laws of logic may derive their existence from linguistic definitions, or from the most general features of reality, or from the structure of ideally rational belief systems, or from something else. The issue is irrelevant because I am concerned, as I said, only with the epistemology of normative theories, not with the ontology of normative laws – only with how we judge the credentials of normative theories, not with what those theories are ultimately *about*. In fact, I do not believe that logical truth is dependent at any point on what human beings actually do: the laws of logic, whatever they may be, hold good irrespective of whether or not there is life in the universe. So it is easy enough to see how a systematically sound theory about the laws of logic might accord with the overwhelming majority of normative judgments that have actually been made and still not accord with all of them. Since it is not being claimed that these laws derive their validity from human behaviour, there is nothing inconsistent in admitting that human inferences sometimes fail to conform to them.

7. **Kahn & Rips*** remark that many idealisations in natural science describe an ideal state in such a way that departures from the ideal may be explained in terms of factors predictable by parameters of the model, whereas nothing like this is the case in regard to normative theories of deducibility or probability. The point is well taken. But other idealisations in natural science do require deviations to be independently explained. If this were not so, the physiologists who tell us how a healthy body functions (and who is perfectly healthy for long?) would already have solved all the problems of pathology.

8. Several commentators (**Daniels & Smith,*** **Lycan,** and **Stich**) object that wide reflective equilibrium would provide a better established normative theory than the relatively narrow equilibrium that I endorse. But the issues that have to be settled in order to transform a narrow reflective equilibrium about deducibility or probability into a wide one tend to have two features that render a quest for the latter inappropriate here. The first feature is that all such issues (about the law of the excluded middle, about deducibility from self-contradictory propositions, about extensionality, and so on) are highly controversial and therefore rather unsuited to being raised in the context of providing background assumptions for the experimental study of irrationality. And the second relevant feature of these issues is that they are all rather fine-grained and therefore intrinsically unsuited to form the topic of experimental tests that presuppose an absence of prior tutoring in the problems and terminology of normative philosophy. **Nisbett** objects that, since ideal conditions for human reasoning would on my view include appropriate education, I ought to admit the

relevance of training in sophisticated normative theories: hence, since these theories are the product of wide reflective equilibrium, it is the latter, he claims, that ultimately underwrites the appropriate criteria for evaluating lay reasoning. But the education to which I referred in the target article (part II, section 2) was education in the incontestable mathematical implications of accepted ideas. That is how a knowledge of Bernoulli's theorem, for example, might be regarded as forming part, under ideal conditions, of a lay individual's competence. It is implicit in his narrow reflective equilibrium. And that is quite a different matter from being trained to approach all logical problems with a Quinean extensionalism, or a Kripkean essentialism, or from being trained to approach all natural-scientific problems with a Popperian horror, or a Carnapian love, of inductive reasoning. Yet it is just such a training that a wider reflective equilibrium might endorse.

9. Finally, a number of commentators argue against my example of how we need to appeal to particular intuitions in order to validate criteria of deducibility. **Grandy**° claims that, for it to be false according to most intuitions that

4. Either, if John's automobile is a Mini, John is rich, or, if John's automobile is a Rolls, John is poor,

it would have to be true that

5. John's automobile is a Mini and John is not rich and John's automobile is a Rolls and John is not poor.

But I am willing to wager Grandy £5 that his claim is experimentally refutable. If he wins, I shall just make all these sentences, (1) through (5), a little more explicit by inserting the word "sole" before each occurrence of "automobile" and shall then renew the wager. **Margalit & Bar-Hillel** claim that (4) here does indeed follow from

3. If John's automobile is a Mini, John is poor, and if John's automobile is a Rolls, John is rich.

But I can reply only that at current U.K. prices (3) is true, for most U.K. Johns, and (4) is false.

III. The competence-performance distinction in regard to deductive and probabilistic reasoning. I argued that the attribution of a correct deductive and probabilistic competence to normal adults is implicit in the epistemological truism that conformity to normal intuitions is a necessary credential for theories about criteria of everyday deducibility and probability. Seven kinds of objection are urged against this attribution.

1. **Einhorn & Hogarth** object that the only reason I give for assuming rationality of competence is "a naive evolutionary adaptationist position." On the contrary. I specifically denied (in part I, section 4 of the target article) that evolutionary considerations are fitted "to predict the precise level of rationality that is required for this or that species' continued survival within its present environment."[1] Einhorn and Hogarth's objection is cogent against what **Lycan**'s commentary says about nature, not against what I say. Instead, the only reason I accept for assuming rationality of competence is the argument relating to intuitions.

2. **Sampson** claims that my arguments "rely heavily on analogies with Chomskian linguistics" and that "the aspects of linguistic methodology appealed to are questionable"; and **Shafer**° makes a similar objection. But in fact all references to analogies with grammar could be removed from my essay and still leave the validity of the argument for rational competence unaffected. So whatever equivocations or other weaknesses may exist in Chomskian linguistics cannot undermine the hard core of the argument in my essay. That argument has to be judged on its own merits. In particular, critics who dislike the argument's reliance on the need for relevant normative theories to conform with dominant intuitions are invited to explain how else they think that the relevance of these theories can be ensured: that is, how else can we be sure that the normative criteria in question are those that apply to the particular concepts of conditionality, disjunction, probability, evidence, relevance, and so on with which lay subjects are actually operating?

3. **Glucksberg** raises an interesting problem about children. If education can help remove inequalities of reasoning performance among adults, like those in regard to Bernoulli's theorem, it seems plausible to hold that it may also help remove such inequalities between children and adults, as the children grow older; and if this is so it looks as though I must have been wrong to say that we are free to ascribe an inferior competence to children. But what I say about children does not depend at all on any empirical theory about intellectual maturation. It depends instead entirely on the fact that children's intuitions are not normally invoked as evidence for the correctness of the criteria for deducibility or probability that are applied to adults' reasoning. It is this that leaves psychologists free to test hypotheses about the correctness of a child's competence, at various ages and under various environmental and educational conditions, whereas the correctness of adult competence is not an issue for experimental enquiry.

4. **Kahneman** complains that to suggest reading off a theory of deductive competence, say, from a theory of deducibility is to make a "disconcerting" proposal for "a new domain of study," and he is worried about who will undertake this work. But no one who is familiar with the relevant literature need be disconcerted here. The work on natural deduction that has been carried out by logicians since 1934 (Quine 1952, pp. 166 ff.) has always aimed at closeness to natural procedures (see Kneale & Kneale 1962, pp. 538 ff.); and the actual reading off of a theory of competence is an utterly trivial procedure involving, at its simplest, the replacement of "may" by "can" in rules of permissible deduction. For example, if "p may be inferred from p & q" were a suitable logical rule (and perhaps primitive in the system of logical theory), then "p can be inferred from p & q" would be a suitable statement of the corresponding competence (and perhaps axiomatic in the theory of competence). Analogously, there has been much discussion among philosophers for half a century or so about which systematic interpretation or interpretations of the

mathematical calculus of chance come closest to the conceptions of probability that operate in everyday life (see Nagel 1939 and Mackie 1973, for summaries and references).

5. I quite agree with **Lopes** that the ascription of a competence cannot explain the performances that accord with it, and that an explanation of these performances is just as much needed as an explanation of the deviant ones. But I was not attempting to provide either kind of explanation in specific terms: I was discussing what needs to be explained rather than how it might be explained.

6. Several commentators (**M. S. Cohen**, **Kahn & Rips**,* **Lycan**, and **Sampson**) object that my theoretical framework allows no room for systematic errors that result from the presence of "satisficing" mechanisms. This is certainly an important issue, on which I have taken a distinctive line. On my view, if a normal person's competence includes some heuristic short-cuts or computational dodges, then these are all intrinsically valid procedures (if correctly operated), like the so-called anchoring and adjustment heuristic – not intrinsically invalid ones, like the "availability heuristic." Where systematic errors occur we have thus to conceive their causation as being mediated through the malfunctioning of otherwise adequate mechanisms rather than through the normal functioning of intrinsically inadequate ones. But I cannot see how any psychological experiments could ever arbitrate between my kind of hypothesis and that of the commentators. **M. S. Cohen** suggests that experimenters might enquire whether evidence for the use of intrinsically invalid procedures might increase with the presence of a disturbing factor (such as fatigue or lack of motivation), independently of any increase in errors. If it does, he says, there would be experimental support for the claim that "heuristics" replace more normatively appropriate procedures under conditions of stress. Not so. If there were no increase in errors, the evidence could just as well be construed as showing the increased use, under a certain kind of stress, of some intrinsically valid dodge, which could be described by the addition of some necessary qualification or restriction to the description of the alleged intrinsically invalid dodge. For example, such a modification of the so-called availability heuristic might take the form: "Take as evidence the data that come most readily to mind unless you have reason to suppose that they may constitute a biased sample." Of course, a further explanation would then have to be found for errors in the operation of this heuristic, for example, errors that manifest neglect of the restriction in regard to biased samples. Moreover, even if we adopted M. S. Cohen's view, we would still have to look, in lieu of such an explanation, for additional factors (or intensifications of existing ones) that promote use of the imperfect heuristic in situations in which it does result in erroneous conclusions. Nor would any short-run failure to discover such a factor count as evidence that it does not exist. Biologists have long since given up supposing that a process is spontaneous just because its cause is hard to discover.

In sum, the issue in dispute here is a purely conceptual one: all possible psychological data could be explained either way. My rejection of the view supported by **M. S. Cohen** et al. stems only from the desire for a coherent conceptual framework that acknowledges the essential role of lay intuitions in the establishment of criteria for evaluating lay reasoning. Some competence for detecting biased selections of evidence, for example, is inherent in the intuitions that have to be invoked as a basis for inductive norms. The strawberries on top of the punnet, we all know, may be rather better than those underneath.

7. **Griggs** is quite right to insist that a purely descriptive and explanatory account of human reasoning – an account that eschews all judgments on validity or fallacy – would not be subject to my argument. Only when researchers seek to determine the conditions under which people reason correctly or incorrectly does a competence-performance distinction come to be needed, and only then can questions arise whether "satisficing" heuristics are in operation. Griggs is on rather weaker ground, however, as the present commentaries help to show, in asserting that "most reasoning researchers are not concerned with experimentally testing for rationality." In any case, the validity or invalidity of an inference is hardly an unimportant feature of it, with respect to the resultant impact on ordinary human concerns. An experimental science of reasoning that was not concerned at all with the extent and conditions of rationality could not hope to say as much to educators, administrators, businessmen, and other potential users of psychological discoveries. Nor would the problem of rationality just disappear if psychology departments ceased to concern themselves with such issues. The problem is so important that it would immediately be taken up by others.

IV. Some issues in the literature. The special problems of normative enquiries like ethics or logic fall outside the accepted field of experimental psychology. Hence experimental psychologists who work on problems about rationality naturally tend to take their normative assumptions at second hand from standard authorities and to apply them in a relatively unselfcritical fashion, while devoting their primary attention to the construction and interpretation of experiments. My target article took an opposite course. I assumed the robustness of certain reported results, which I took at second hand from the literature, and tried to explore two main issues about normative criteria of rationality: first, what is presupposed by invoking them and, second, how do such criteria relate to the experimental results? I have already discussed the points raised by commentators in regard to my remarks on the first of these two issues, and I turn now to comments about my remarks on the second. These comments were concentrated around seven areas of experimental enquiry, which I shall discuss in the same order as in the original essay.

1. *The four-card task.* **Weisberg** objects that my content-based explanation of performance error in this task commits me to a noncompetence view, in the sense that I seem to be arguing against an abstract system of rules. Similarly **Evans & Pollard** object that my definition of competence becomes unclear if it is held to be situation specific. But there is a misunderstanding here, for which my own phraseology is largely to blame.

Instead of talking about people's failure "to apply their logical competence" to certain tasks because their competence is not formulated expressly at a sufficient level of generality, I should have written instead of their failure to apply the mechanism that underlies their competence to certain tasks – a failure that is facilitated by the fact that no sufficiently general formulation of the principle embodied in this mechanism has ever risen to the level of consciousness. So the competence in question, namely the ability to contrapose, is a quite general one, the nonexercise of which in certain tasks demands an explanation. And the matching bias which seems to be operating here may well be favoured, as **Wetherick** suggests, by a mistaken analogy with the "All ravens are black" kind of situation, where it is normally a mistake to contrapose for purposes of testing (see Cohen 1970, pp. 97ff.).

Wason's main criticism of my treatment of the four-card phenomenon is that I assert that the subjects are indulging in a form of reasoning that on a few moments' prompted reflection they would be willing to admit to be invalid, whereas in fact "a fair proportion of subjects" fail to correct their initial responses even when all the relevant information is made available to them. However, my claim is just that the subjects are capable of recognising their error here, when it is properly pointed out to them (and the prompting may need to be pretty explicit). My claim is not that they won't make the same mistake again. In order to justify description of this phenomenon as a cognitive illusion – a description that Wason endorses – one must be able to preserve enough of the analogy with those perceptual errors (the apparently bent stick in water, the mirage on the road, the Müller-Lyer, conjuring tricks, and so on) that are paradigmatic for the concept of illusion. Now the position with visual illusions is that people very often go on seeing things in the wrong way, even when they have accepted (as a result of feeling the stick, coming up to the mirage, measuring the lines) that they were mistaken on the previous occasion. That is why over many centuries such illusions afforded Pyrrho, Sextus Empiricus, Montaigne, and others with such plausible arguments for scepticism. Accordingly, the four-card phenomenon would not justify description as an illusion unless its sources were so deep that the tendency to make the relevant mistake resisted easy eradication in many people. Of course, if Wason could demonstrate that a high percentage of subjects cannot even understand, within a relatively short period, that they have made a mistake when it is carefully pointed out to them why they have done so, he would certainly show that he and I are wrong to call the four-card phenomenon an illusion. But the papers that he cites include no such demonstration.

2. *Availability.* On the availability issue **Weisberg** reiterates Tversky and Kahneman's explanation in terms of an intrinsically invalid heuristic, and implies that this is a simpler type of explanation than the one I offer. But such a difference in simplicity is not at all obvious. On Weisberg's account the underlying mechanism raises a query about representativeness, to which the alleged heuristic permits ease of recall to supply an affirmative answer. On my account the underlying mechanism raises the same question, and the failure to answer it is overlooked because of the immediacy of the data. By what relevant criteria is the former process simpler than the latter?

3. *Sample size and regression.* In relation to the commission of fallacies by "the world's experts in the field of mathematical psychology," **Sternberg** claims that *their* fallacies cannot result from inadequate training. Why not? It is surely not to be supposed that in so young a subject as statistical theory, where the foundations are still highly controversial, pedagogic methods have already been perfected beyond the possibility of any improvement. Indeed there seems no practicable way of establishing, at any one time, that the pedagogics of statistical theory do not admit of further improvement in efficiency. No doubt someone may object that I am therefore advancing an empirically unfalsifiable claim. Yes indeed, that is what I am doing, though only in the sense in which "Everything has some explanation" is unfalsifiable. Here too, as in what was said above about the existence of causes for the malfunction of intrinsically valid heuristic mechanisms, my primary purpose is to sketch a coherent framework of enquiry, not to propose conjectures within such a framework. Specifically, the failure of some experts to live up to expectations should direct attention towards finding ways to improve the methods for training and selecting them: we should not sit back complacently and suppose that, if even they make mistakes of a certain kind, then the tendency to do so must somehow be an ineradicable feature of the human mind. And, if some conjecture is called for, my guess is that in such contexts a little gentle mockery by colleagues would be a very effective reminder of lessons that have been forgotten. On the other hand, suppose it *can* be shown (as Tversky and Kahneman do not attempt to show) that when a fallacy of the kind in question is committed by an expert, he will still tend to commit it again on later occasions even after accepting that he was in error on the first occasion. All that is thereby shown is that the fallacy has to be regarded as a rather sophisticated form of illusion.

In this connection **Tversky** makes the rather odd comment that "the psychologist is very interested in whether naive subjects have learned from lifelong experience that nonrepresentative results are more frequent in small than in large samples." The first thing that is puzzling here is that the subjects cited do not seem to have been all septuagenarians or octogenarians, and Bar-Hillel's (1979) subjects, for example, were all either students or applicants for university entry: how could such subjects be expected to give a fair indication of what people "have learned from lifelong experience"? Alternatively, if Tversky means by "lifelong" just "in their lives to date," then it seems strange that psychologists have not studied whether increasing age has any effect on the relevant learning. Second, what Bernoulli proved – the law about sample size that is relevant to the estimation of statistical magnitudes – is a mathematical truth about the classical concept of probability. This truth holds independently of anything that happens in the perceivable world. In particular, it is independent of the empirical question whether, over all samples actually selected in some way from physically real populations (finite and

infinite), nonrepresentative results are more frequent in small than in large samples. Tversky implies that he knows the latter question to require an affirmative answer. The rest of us can only await the publication of his research in the matter with an interest tinged ever so slightly with the politest of scepticisms.

Even if **Tversky** has found, by judicious selection and exhaustive sample extraction, some actual populations and some actual samples for which his bold empirical hypothesis holds good and none for which it does not, this would scarcely justify extrapolation to populations of very different kinds that are too large or open ended for probabilities in them to be determined independently of a partial sampling procedure. Nor can there be any serious reason to suppose that naive subjects, without a misguided zeal for putting mathematical laws to empirical tests, would ever deliberately indulge in – let alone happen accidentally upon – the combinatorial explosions that are required for the exhaustive extraction of different-sized samples from relatively large populations.

4. *The fallacy of the converse.* I am happy to find that, as **Revlin**° points out, my remarks about this fallacy have been partly anticipated. However, those remarks were addressed only to the alleged prevalence of the fallacy: I was not denying that it is ever committed. **Weisberg** assumes that a general human tendency towards illicit conversion has indeed been experimentally established, and suggests that it results from subjects' ignorance of how to carry out computations in truth-functional logic. But he does not answer my argument, or the arguments of others, that no such tendency has been established and so there is no such tendency to explain. Nor does he answer the argument (in part I, section 1 of the essay) that the logic of natural language is not truth functional.

5. *The gambler's fallacy.* The gambler's fallacy "is as current today as it was in Bernoulli's time," **Diaconis & Freedman** claim. This is another bold empirical claim, and again – on the face of it – not an easy one to substantiate. I look forward to the early publication of Diaconis & Freedman's evidence (or to their avowal of having none). Meanwhile I can only emphasise that, despite their (admittedly "brutal") reformulation of what I wrote, I was not denying that anyone ever reasons fallaciously to the maturity of a chance. I was merely pointing out (and this is my answer also to **Skyrms's** query) that at least three alternative interpretations need to be excluded or discounted in any case in which it is suspected that someone might be so reasoning, and that more empirical work needs to be done on the subject. That is quite different from "elevat[ing] the notorious gambler's fallacy," which is what **Margalit & Bar-Hillel** accuse me of; and their "alarm" at "Cohen's inclination not only to excuse, but to actually condone, people's mistakes" shows a certain disposition to shy at shadows.

6. *The cabs.* The essence of my claim here is that, since the probability at issue is the Pascalian[2] unconditional probability that the cab involved in the accident was blue, information predictive of relevant base rates would be information about the relative frequencies of blue and green cabs among cabs recently involved in accidents of the same kind. I am not trying to establish a Baconian analysis of the problem (**Niiniluoto**), nor am I repudiating Bayes's theorem (**Levi, Shafer**°). What I actually suggested was that in the absence of information determining appropriate prior probabilities here the subjects did right to "suppose equal predispositions" (as I put it), in accordance with what Niiniluoto, in his valuable discussion of the history of the problem (which effectively refutes what **Diaconis & Freedman** say about Todhunter's [1949] lottery), calls the traditional strategy in the Bayes-Laplace school. It follows that if there are no cabs of one colour at all, and all have the other colour (a possibility raised by **Margalit & Bar-Hillel** and **Skyrms**), then a fortiori there is a zero base rate for cabs of that colour being involved in accidents of the relevant kind. But the subjects were just told of an 85%–15% distribution in cab colour, and this fact, or any other distribution that allows some cabs of either colour, is a very weak foundation for an estimate of the *relevant* base rate. It is certainly quite surprising that not only a few psychologists, like **Tversky** and **Kahneman,** but even a few people who say that they "make a living by teaching probability and statistics," like Diaconis & Freedman, should persist in supposing otherwise, even after their error has been pointed out. Specifically, why on earth should it be supposed that subjects, asked to estimate the unconditional probability that the cab *involved in the accident* was blue, ought to take into account a prior distribution of colours that would at best be relevant only if the issue at stake was just about the colour of a cab that was said to have been seen somewhere, *not necessarily in an accident,* and was taken to be blue? Until Tversky, or Kahneman, or one of their apologists, supplies a satisfactory answer to this question, the scientific world will not be much impressed by their response to those who criticise them.[3]

Analogous considerations apply to the diagnosis example. Bilharzia is now one of the commonest diseases in the world, but it would be rather absurd to take its current frequency in the world population as predictive of the base rate, or prior probability, when diagnosing a patient who never wades in fresh water. Those (**Blackburn, Evans & Pollard, Kahneman, Krantz, Mackie, Margalit & Bar-Hillel, Sternberg, Zabell**) who hold that gross epidemiological statistics are relevant here, irrespective of the patient's own susceptibilities, are refusing to take into account what Keynes (1921) called the "weight" of evidence.

The conclusion desired here is an unconditional probability about a single case, p(A), and that conclusion has to be detached from a conditional probability, p(A | E). People sometimes think, as **Mackie** evidently does, that such a detachment is legitimate just so long as E includes all our relevant knowledge. But wherever our relevant knowledge is rather limited in extent, and so the weight of the evidence is low, this policy can lead to disastrously bad estimates of the unconditional probability. Any insurance company that adopted such a policy would rapidly become bankrupt. Instead, we need to increase the weight of the evidence by discovering quite a substantial amount of the causally relevant facts (which means taking into account the

causally relevant susceptibilities of the patient or the specific circumstances of the cab sighting) and determining the corresponding conditional probability, before we detach the unconditional probability. Of course, it may well be impossible to be perfect here. But that is no excuse for detaching the unconditional probability on the basis of excessively lightweight evidence. Moreover, just as some kinds of fact (such as the patient's life-style or his past medical history) are weight-increasing with regard to the detachment of a particular unconditional probability, so too, other kinds of fact may be weight-reducing.

Suppose, for example, that your name is Algernon Charles Thomas and that, though nothing is known about the frequencies of diseases A and B in the world, in your country, or in your city, it does happen to be known that among the twelve other (all unrelated) Algernon Charles Thomases, scattered over different continents and cultures, disease A is at the moment twice as common as disease B. To compound a base-rate probability estimated from that accidental and very weakly predictive statistic, with the probability deriving from the diagnostic test, would obviously be weight-reducing. No doubt it is a matter for judgment, experience, and expertise in particular cases to distinguish weight-increasing from weight-reducing evidence. My claim was just that it is not unreasonable for subjects to suppose that the distribution of cab colours in a particular city, or current world-wide epidemiological statistics, are relatively nonpredictive, and generate weight-reducing base rates, in cases of the kinds in question. Moreover, if I am right here, then where no particular cause of error is operating, one would expect subjects with a competence for rational inference to take positive account of base-rate statistics about causally relevant, or relevantly specific, features. And this is just what Bar-Hillel (1980) has shown that they in fact do (though unfortunately she takes such a procedure to be fallacious).

Of course, the concept of weight is open to alternative explications. **Levi** gives a very clear account of how Tversky and Kahneman's error may be exposed on a subjectivist analysis of probability judgments. I myself prefer to suggest that the difference between weight-increasing and weight-reducing evidence, in respect of a particular outcome, corresponds exactly with the difference between evidence capable of determining propensity-type Pascalian probabilities, and evidence capable only of determining short-run relative frequencies, for that kind of outcome. But the issue between Levi's account and mine is not material here.[4] What is important is just that sufficiently scrupulous attention be paid to the various alternative semantics that are available for the calculus of chance – alternative theories of what is meant by statements about "probability" – so that equivocations and other sources of confusion are avoided. Several of the commentators, however, do not show any obvious awareness of the vast literature that exists on this subject. For example, **Tversky** says that I "[introduce] a new concept of propensity probability which . . . satisfies the classical probability axioms." But in fact I actually gave references to recent discussions of this concept by Popper (1959b; 1968) and by Mellor (1971); and Hacking (1975) traces such a theory of probability back to the seventeenth century.

It is **Diaconis & Freedman** who do come out with a striking new interpretation for the classical calculus. They tell us, in a footnote, that "p(A | B) can be interpreted as the probability of A, if you know that B has occurred." But this unfortunately confuses "p(A | B)" with "p(A | you know that B has occurred)," and "p(A | B) = n" both with "if you know that B has occurred, then p(A) = n" and with "p(if you know that B has occurred, then A) = n," as well as muddying several other familiar issues. Unless **Tversky** and **Kahneman** and their apologists are prepared to pay adequate attention here to the relevant literature (which also has quite a lot to say about the difference between empirical and mathematical issues), they will continue to risk making further errors of the kind that most naive subjects may well avoid.

However, I think that my remarks about the disease case need amplification in one respect, in order to dissolve the paradox that **Mackie** describes. When I wrote that "the literature under criticism is propagating an analysis that could increase the number of deaths from a rare disease," I meant, as the subsequent paragraph was intended to show, that this increase would take place if, as a result of such papers' being published, less regard were paid to the *weight* of evidence than is currently paid. Specifically, this result would follow if physicians supposed that the current worldwide frequency of a disease was to be taken into account and treated as predictive rather than its current frequency among patients having this or that kind of susceptibility. An administrator of a hospital that took in patients at random from all over the world might still secure good results if he insisted on all those with the symptoms in question being diagnosed as having the much commoner disease. But even better results could be secured, though at greater expense, by increasing the weight of the evidence on which diagnoses were based (see further Cohen 1980a, especially p. 60).

7. *Representativeness*. Since I still claim in part II, section 4 of the target article that a Baconian analysis makes good sense of judging probability in terms of representativeness, I am at a loss to understand why **Tversky** supposes that I have "abandoned" the Baconian theory. What was proposed quite explicitly in Cohen (1979 p. 397) was that experimenters should be prepared for their subjects to use different conceptions and criteria of probability in different contexts, and that it was worth while seeking cues to the discovery of which conception is used when. I have in fact always rejected the "either-or" kind of dogmatism that Tversky imputes to me. Just as people measure fruit sometimes by weight, sometimes by number, and sometimes by volume, according to what seems most appropriate in the context; just as they judge the goodness of a play by criteria different from those by which they judge the goodness of a car; so too we should not be surprised if not only several different Pascalian conceptions of probability may be found in common use but also at least one non-Pascalian conception.[5] Even Bernoulli, as Shafer (1978) has now conclusively shown, was prepared to investigate the mathe-

matics of nonadditive probabilities: He did not suppose the complementational law for negation to be quasi God-given,[6] as Tversky seems to do.

Yet another of **Tversky's** easily verifiable misconstruals of what I say is his claim that "Cohen actually assigns positive inductive probability to self-contradictory propositions." What I actually assign to these propositions is zero inductive probability (Cohen 1977b, p. 238, theorem 614). Here, as elsewhere (see Cohen 1979 pp. 403–405) Tversky confuses conditional and unconditional probabilities. In the Baconian system p(A & not-A | B) may have a positive value in certain circumstances, but p(A & not-A) is always zero. This property of conditional probabilities in the Baconian system, along with their contraposability, makes possible a generalisation of argument by reductio ad absurdum. If a contradiction is deducible from a set of premises then the premises are *demonstrably* not cotenable: correspondingly, if a contradiction is rendered inductively probable by a set of premises, then the premises are *probably* not cotenable.)

Tversky also complains that in the Baconian system at most one of two inconsistent outcomes can have a positive probability. But that is just because this kind of probability judgment states where the evidence is pointing (and how weighty it is). If the totality of relevant facts points in one direction, it can't point in the other direction also. To complain about this is like complaining that in a Pascalian system we can't allow the same evidence to give greater than prior probability to each of two inconsistent outcomes.

ACKNOWLEDGMENT

I am grateful to Mrs. D. Faulkner for valuable research assistance.

NOTES

1. To that extent I now (hereby and herewith) repudiate the quip that **Einhorn & Hogarth** correctly quote against me from an earlier paper (Cohen 1980d).

2. In calling a probability function "Pascalian" I mean that it conforms to the principles of the classical calculus of chance – that is, to the standard mathematics of probability.

3. Tversky complains also that I do not discuss the relations between propensity probabilities and Baconian ones. But that topic is discussed in Cohen (1977b), to which I referred in the target article.

4. Restrictions on the applicability of a subjective analysis to probability judgments about past events are discussed in Cohen (1977b, pp. 89–91) and in Cohen (forthcoming).

5. There seem in fact to be at least two possible kinds of probability interpretation for the formal system underlying what I call Baconian probabilities. One is the objective interpretation (related to causality) which I prefer: the other is the subjective interpretation constituted by Shackle's (1949) and Levi's (1967) theory of potential surprise (see Cohen 1980c, p. 171).

6. In the last paragraph of his commentary **Tversky** implies that he has evidence that his subjects admit to having been in error when they judged probabilities by representativeness. When this new evidence is published, it will need to exclude the possibility that the circumstances of the experiment are such as to make naive subjects suppose that they were asked originally for a judgment of Pascalian probability. There is a real problem about indeterminacy here (see Cohen 1979, p. 398).

EDITORIAL NOTE

°Asterisks indicate commentary will appear in forthcoming Continuing Commentary.

References

Anderson, A. R. & Belnap, B. D. (1974) *Entailment: the logic of relevance and necessity*. Vol. 1. Princeton: Princeton University Press. [LJCar]

Anderson, N. H. & Shanteau, J. C. (1970) Information integration in risky decision making. *Journal of Experimental Psychology* 84:441-45. [LLL]

Armstrong, D. M. (1973) *Belief, truth, and knowledge*. Cambridge: Cambridge University Press. [WGL]

Ayer, A. J. (1946) *Language, truth and logic*, London: Gollancz. [LJCa]

Bar-Hillel, M. (1979) The role of sample size in sample evaluation. *Organizational Behavior and Human Performance* 24:245–57. [LJCr, AM]

(1980) The base-rate fallacy in probability judgments. *Acta Psychologica* 44:211–33. [LJCr, AM]

Beach, L. R. & Peterson, C. R. (1966) Subjective probabilities for unions of events. *Psychonomic Science* 5:307–8. [LLL]

Bergson, H. (1903) *Introduction à la métaphysique*. Paris: A. Bourgeois (Cahiers de la quinzaine). [LJCa]

Bernoulli, J. (1713) *Ars conjectandi*. Basle. [LJC]

Berry, J. W. & Dasen, P. R. (1974) History and method in the cross-cultural study of cognition." In: *Culture and cognition: Readings in cross-cultural psychology*, Berry & Dasen. London: Methuen. [SPS]

Billingsley, P. (1979) *Probability and measure*. New York: Wiley. [BS]

Blackburn, S. (1973) *Reason and prediction*, Chs. 5–6, Cambridge: Cambridge University Press. [SB]

(1980) Opinions and chances. In: *Prospects for pragmatism*, ed. D.H. Mellor. Cambridge: Cambridge University Press. [SB]

Brehmer, B. (1980) In one word: Not from experience. *Acta Psychologica* 45:223–41. [HJE]

Brunswik, E. (1956) *Perception and the representative design of experiments*. 2d ed. Berkeley: University of California Press. [HJE]

Carnap, R. (1947) *Meaning and necessity*. Chicago: University of Chicago Press. [LJCa]

(1950) *Logical foundations of probability*. Chicago: University of Chicago Press. [LJCa]

Chapman, L.J. & Chapman, J.P. (1967) Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology* 72(3):193–204. [LJCa]

(1969) Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology* 74(3):271–80 [LJCa]

Chomsky, N. (1965) *Aspects of the theory of syntax*. Cambridge, Mass.: MIT Press. [LJCa]

(1968) *Language and mind*. New York: Harcourt Brace Jovanovich. [SG]

Clifford, B. R. & Bull, R. (1978) *The psychology of person identification*. London: Routledge and Kegan Paul. [SLZ]

Cohen, L. J. (1970) *The implications of induction*. London: Methuen. [LJCar,]

(1971) Some remarks on Grice's views about the logical particles of natural language. In: *Pragmatics of natural language*, ed. Y. Bar-Hillel. Dordrecht: Reidel. [LJCa]

(1977a) Can the conversationalist hypothesis be defended? *Philosophical Studies* 31:81–90. [LJCa]

(1977b) *The probable and the provable*. Oxford: Oxford University Press. [LJCa, JLM, AT]

(1979) On the psychology of prediction: Whose is the fallacy? *Cognition* 7:385–407. [LJCar, IN, AT, RWW]

(1980a) Bayesianism versus Baconianism in the evaluation of medical diagnoses. *British Journal for Philosophy of Science* 31:45–62. [LJCr]

(1980b) Some historical remarks on the Baconian conception of probability. *Journal of the History of Ideas* 41:219–31. [LJCa]

(1980c) What has induction to do with causality? In *Applications of inductive logic*, ed. L. J. Cohen & M. Hesse, Oxford: Clarendon Press. [LJCr]

(1980d) Whose is the fallacy? A rejoinder to Daniel Kahneman and Amos Tversky. *Cognition* 8:89–92. [LJCar, HJE, IN, AT, RWW]

(1982, forthcoming) Intuition, induction and the middle way. *Monist* 65, pt. 3. [LJCr]

(forthcoming) Subjective probability and the paradox of the gatecrasher. *Arizona State Law Journal*. [LJCr]

Cole, M. & Bruner, J. S. (1971) Cultural differences and inferences about psychological processes. *American Psychologist* 26:867–76. [SPS]

Copi, I. M. (1954) *Symbolic logic*. New York: Macmillan. [LJCa]

Daniels, N. (1979) Wide reflective equilibrium and theory acceptance in ethics. *Journal of Philosophy* 76:256–82. [LJCa]

(1980) On some methods of ethics and linguistics. *Philosophical Studies* 37:21–36. [LJCa]

Davidson, D. (1966) The logical form of action sentences. In: *The logic of decision and action*, ed. N. Rescher. Pittsburgh: University of Pittsburgh Press. [LJCa]

Dawes, R. M. & Corrigan, B. (1974) Linear models in decision making. *Psychological Bulletin* 81:95–106. [LJCa]

de Morgan, a. (1856) On the symbols of logic. *Transactions of the Cambridge Philosophical Society* 9:120. [PD]

Dennett, D. C. (1978) Intentional systems. In: *Brainstorms*. Montgomery: Bradford Books. [WGL]

(1979) *Brainstorms*. Hassocks: Harvester. [LJCa]

(forthcoming a) Three kinds of intentional psychology. In a Thyssen Philosophy Group volume to be edited by R. A. Healey. [WGL]

(forthcoming b) True believers: The intentional strategy and why it works. In a volume of 1979 Herbert Spencer lectures. Oxford: Oxford University Press. [WGL]

Dummett, M. (1978) The justification of deduction. In: *Truth and other enigmas*, ed. M. Dummett. London: Duckworth. [LJCa]

Dumont, B. (1980) L'influence du language et du contexte dans des epreuves de type "logique." Ph.D. thesis, University of Paris. [PW)

Edwards, W. (1975) Comment. *Journal of the American Statistical Association* 70:291–93. [LJCa]

Einhorn, H. J. (1980) Learning from experience and suboptimal rules in decision making. In: *Cognitive processes in choice and decision behavior*, ed. T. S. Wallsten, pp. 1–20. Hillsdale, N. J.: Erlbaum. [HJE]

Einhorn, H. J. & Hogarth, R. M. (1978) Confidence in judgment: Persistence of the illusion of validity. *Psychological Review* 85:395–416. [HJE]

(1981) R. M. Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology*, 32:53–88. [LJCa, HJE, LLL]

Ellis, B. (1979) *Rational belief systems*. Oxford: Blackwell. [LJCa]

Evans, J. St. B. T. (1980) Current issues in the psychology of reasoning. *British Journal of Psychology* 71:227–39. [RAG]

(1982, in press) *The psychology of deductive reasoning*. London: Routledge & Kegan Paul. [JStBTE]

Falmagne, R. J., ed. (1975) *Reasoning: Representation and process*. Hillsdale, N. J.: Erlbaum. [RAG]

Finetti, B. de (1931) Sul significato soggettivo della probabilità. *Fundamenta mathematica* 17:298–329. [LJCa]

Finocchiaro, M. (1981) Fallacies and the evaluation of reasoning. *American Philosophical Quarterly* 18:13–22. [LJCr]

Fischhoff, B. (1975) Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance* 1:288–99. [LJCa]

Fodor, J. A. & Garrett, M. (1966) Some reflections on competence and performance. In: *Psycholinguistics Papers*, ed. J. Lyons & R. J. Wales. Edinburgh: Edinburgh University Press. [GS]

Freedman, D.; Pisani R. & Purves, R. (1978) *Statistics*. New York: W. W. Norton. [SLZ]

Glass, A. L.; Holyoak, K. J. & Santa, J. L. (1979) *Cognition*. Reading, Mass.: Addison-Wesley. [RWW]

Goldberg, L.R. (1970) Man versus model of man: A rationale, plus some evidence, for a method of improving on clinical inferences. *Psychological Bulletin* 73(6):422–32. [LJCa]

Goldman, A. (1976) Discrimination and perceptual knowledge. *Journal of Philosophy* 73:771–91. [WGL]

Good, I. J. (1971) Reply to Professor Barnard. In: *Foundations of statistical inference*, ed. V. P. Godambe & D. A. Sprott. Toronto: Holt, Rinehart & Winston. [SLZ]

Goodman, N. (1954) *Fact, fiction, and forecast*. London: Athlone. [LJCa, AM]

(1979) *Fact, fiction, and forecast*. 3rd ed. Indianapolis: Hackett. [SLZ]

Grace, N. D.; Muench, H. & Chalmers, T. C. (1966) The present status of shunts for portal hypertension in cirrhosis. *Journal of Gastroenterology* 50:646–91. [SLZ]

Grice, H. P. (1975) Logic and conversation. In: *The logic of grammar*, ed. D. Davidson and G. Harman. Encino, Cal.: Dickenson. [LJCa, RWW]

Griffin, N. (1977) *Relative identity*. Oxford: Oxford University Press. [LJCa]

Haack, S. (1974) *Deviant logic*. Cambridge: Cambridge University Press. [LJCa]

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

Hamblin, C. L. (1970) *Fallacies*. London. Methuen. [LJCa]

Hammerton, M. (1973) A case of radical probability estimation. *Journal of Experimental Psychology* 101:252–54. [LJCa]

Hammond, K. R. (1955) Probabilistic functionalism and the clinical method. *Psychological Review* 62:255–62. [HJE]

(1978) Toward increasing competence of thought in public policy formation. In: *Judgment and decision in public policy formation*, ed. K. R. Hammond, pp. 11–23. Denver: Westview. [HJE]

Hempel, C. G. (1945) Studies in the logic of confirmation. *Mind* 54:1–26 & 97–121. [NEW]

Henle, M. (1962) On the relation between logic and thinking. *Psychological Review* 69:366–78. [LLL]

(1978) Foreword. In: *Human reasoning*, ed. R. Revlin & R. E. Mayer, pp. xiii–xviii. Washington, D. C.: Winston. [MH]

Hogarth, R. M. (1975) Cognitive processes and the assessment of subjective probability distributions. *Journal of the American Statistical Association* 70:271–89. [LJCa]

Hume, D. (1739) *A treatise of human nature*. London: John Noon. [LJCa, SLZ]

Hume, D. (1975) *Enquiries concerning human understanding and concerning the principles of morals*. Ed. L. A. Selby Bigge. Third Ed. rev. by P. H. Nidditch. Oxford: Oxford University Press. [JLM]

Husserl, E. (1911) Philosophie als strenge Wissenschaft. *Logos* 1:289–341. [LJCa]

Jeffrey, R. C. (1980) How is it reasonable to base preferences on estimates of chance? In: *Science, belief and behaviour*, ed. D. H. Mellor. Cambridge: Cambridge University Press. [LJCa, AM]

Jenkins, H. M. & Ward, W. C. (1965) Judgment of contingency between responses and outcomes. *Psychological Monographs: General and Applied* 79, whole no. 594. [LJCa]

Johnson-Laird, P.N.; Legrenzi, P. & Sonino Legrenzi, M. (1972) Reasoning and a sense of reality. *British Journal of Psychology* 63:395–400. [LJCa, RAG, RWW, NEW]

Johnson-Laird, P.N. & Wason, P.C. (1970) A theoretical analysis of insight into a reasoning task. *Cognitive Psychology* 1:134–48. [LJCa]

(1977) Postscript. In: *Thinking: Readings in Cognitive Science*, ed. P. N. Johnson-Laird & P. C. Wason. Cambridge: Cambridge University Press. [LJCa]

Kahneman, D. & Tversky, A. (1972a) On the psychology of prediction. *Oregon Research Institute Research Bulletin* 12, whole no. 4. [LJCa]

(1972b) Subjective probability: A judgment of representativeness. *Cognitive Psychology* 3: 430–54. [LJCa, RWW]

(1973) On the psychology of prediction. *Psychological Review* 80:237–51. [LJCa, RWW]

(1974) Subjective probability: A judgment of representativeness. In: *The concept of probability in psychological experiments*, ed. C. A. S. Stael von Holstein. Dordrecht: Reidel. [LJCa]

(1979) On the interpretation of intuitive probability: A reply to Jonathan Cohen. *Cognition* 7:409–11. [LJCar, AT, RWW]

(1981, in press) On the study of statistical intuitions. In: *Judgment under uncertainty: Heuristics and biases*, ed. D. Kahneman, P. Slovic & A. Tversky. Cambridge University Press. [AT]

Keynes, J. M. (1921) *A treatise on probability*. London: Macmillan. [LJCr]

Klotz, I. (1980) The n-ray affair. *Scientific American* 242:168–75. [SLZ]

Kneale, W. & Kneale, M. (1962) *The development of logic*. Oxford: Clarendon Press. [LJCr]

Kocher, P. H. (1957) Francis Bacon on the science of jurisprudence. *Journal of the History of Ideas* 18:3–26. [LJCa]

Kolmogorov, (1950) A. *Foundations of probability*. New York: Chelsea. [LJCa]

Kripke, S. (1972) Naming and necessity. In: *Semantics of natural language*, ed. D. Davidson and G. Harman, Dordrecht: Reidel. [LJCa]

Kyburg, H. E. (1974) *The logical foundations of statistical inference*. Dordrecht: Reidel. [IL]

Lakatos, I. (1970) Falsificationism and the methodology of scientific research programs. In: *Criticism and the growth of knowledge*, ed. I. Lakatos and A. E. Musgrave. Amsterdam: North Holland. [LJCa]

Lemmon, E. J. (1965) *Beginning logic*. London: Nelson. [LJCa]

Levi, I. (1967) *Gambling with truth*. New York: Knopf. [LJCr]

(1977) Direct inference. *Journal of Philosophy* 74:5–29. [IL]

(1980) *The enterprise of knowledge*. Cambridge, Mass.: MIT Press. [IL]

Lewis, C. I. & Langford, C. H. (1959) *Symbolic logic*. New York: Dover. [LJCar]

Lewis, D. (1973) *Counterfactuals*. Oxford: Blackwell. [LJCa]

Lewontin, R. C. (1978) Adaptation. *Scientific American* 239:212–30. [HJE]

(1979) Sociobiology as an adaptationist program. *Behavioral Science* 24:5–14. [HJE]

Lichtenstein, S. & Slovic, P. (1973) Response-induced reversals of preference in gambling: An extended replication in Las Vegas. *Journal of Experimental Psychology* 101:16–20. [LJCa]

Loftus, E. F. (1979) *Eyewitness testimony*. Cambridge, Mass.: Harvard University Press. [SLZ]

Lopes, L. L. (1976) Model based decision and inference in stud poker. *Journal of Experimental Psychology: General* 105:217–39. [LLL]

(1980) Doing the impossible: A note on induction and the experience of randomness. Paper presented at 18th Bayesian Research Conference, Los Angeles, Cal. [LJCa]

Lopes, L. L. & Ekberg, P. H. S. (1980) Test of an ordering hypothesis of risky decision making. *Acta Psychologica* 45:161–67. [LLL]

Lyon, D. & Slovic, P. (1976) Dominance of accuracy information and neglect of base rates in probability estimation. *Acta Psychologica* 40:287–98. [LJCa, AM]

Mackie, J. L. (1973) *Truth, probability, and paradox*. Oxford: Oxford University Press. [LJCr, JLM]

(1974) *The cement of the universe: A study of causation*. Oxford: Oxford University Press. [JLM]

Manktelow, K. I. & Evans, J. St. B. T. (1979) Facilitation of reasoning by realism: Effect or non-effect. *British Journal of Psychology* 70:477–88. [LJCa, RAG, NEW]

March, J. G. (1978) Bounded rationality, ambiguity, and the engineering of choice. *Bell Journal of Economics* 9:587–608. [LLL]

Mellor, D. H. (1971) *The matter of chance*. Cambridge: Cambridge University Press. [LJCar]

Meyer, D. E. (1980) Some/is: An investigation of logical thinking. Ph.D. dissertation, Graduate Faculty, New School for Social Research. [MH]

Mill, J. S. (1874) *A system of logic*. 8th ed. New York: Harper. [MH]

Mises, R. von. (1957) *Probability, statistics and truth*. London: Allen & Unwin. [LJCa]

Moravcsik, J. M. E. (1969) Competence, creativity, and innateness. *Philosophy Forum* 1:407–37. [GS]

Münsterberg, H. (1908) *On the witness stand; Essays on psychology and crime*. New York: The McClure Company. [SLZ]

Mynatt, C. R.; Doherty, M. E. & Tweney, R. D. (1977) Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology* 29:85–95. [LJCa]

(1978) Consequences of confirmation and disconfirmation in a simulated research environment. *Quarterly Journal of Experimental Psychology* 30:395–406. [LJCa]

Nagel, E. (1939) Principles of the theory of probability. *International Encyclopaedia of Unified Science*, vol. 1, no. 6. Chicago: University of Chicago Press. [LJCar]

Neisser, U. (1976) *Cognition and reality*. San Francisco: W. H. Freeman. [RAG]

Newell, A. (1980) Reasoning, problem solving and decision processes: The problem space as a fundamental category. In: *Attention and performance VIII*, ed. R. Nickerson. Hillsdale, N. J.: Erlbaum. [RAG]

Nisbett, R. E. & Borgida, E. (1975) Attribution and the psychology of prediction. *Journal of Personal and Social Psychology* 32:932–43. [LJCa]

Nisbett, R. E. & Ross, L. (1980) *Human inference: Strategies and shortcomings and social judgment*. Englewood Cliffs: Prentice-Hall. [LJCar, JStBTE, SPS]

Piaget, J. (1972) *The principles of genetic epistemology*. London: Routledge and Kegan Paul. [LJCa]

Popper, K. R. (1959a) *The logic of scientific discovery*. London: Hutchinson. [LJCar]

(1959b) The propensity interpretation of probability. *British Journal for Philosophy of Science* 10:25–42. [LJCar]

(1968) On the rules of detachment and so-called inductive logic. In: *The problem of inductive logic*, ed. I. Lakatos. Amsterdam: North Holland. [LJCar]

Postman, L. & Tolman, E. C. (1959) Brunswik's probabilistic functionalism. In: *Psychology: A study of a science*, vol. 1, ed. S. Koch, pp. 502–64. New York: McGraw-Hill. [HJE]

Quine, W. V. O. (1952) *Methods of logic*. London: Routledge and Kegan Paul. [LJCar]

(1960) *Word and object*. Cambridge, Mass.: MIT Press. [LJCa]

Ramsey, F. P. (1931) *The foundations of mathematics*. London: Routledge and Kegan Paul. [LJCar]

Randi, J. (1975) *The magic of Uri Geller*. New York: Ballantine Books. [SLZ]

Rawls, J. (1972) *A theory of justice*. Oxford: Oxford University Press. [LJCa]

Reich, P. A. (1969) The finiteness of natural language. *Language* 45:831–43. [GS]

Reichenbach, H. (1944) *Philosophic foundations of quantum mechanics*. Berkeley: University of California Press. [LJCa]

(1949)*The theory of probability*. Berkeley: University of California Press. [LJCa]

Reid, T. (1969) *Essays on the intellectual powers of man*. Cambridge, Mass.: MIT Press, reproduced from 1814–15 ed., first published 1785. [LJCr]

Revlin, R., & Leirer, V. O. (1978) The effect of personal biases on syllogistic reasoning: Rational decisions from personalized representations. In: *Human reasoning*, ed. R. Revlin & R. E. Mayer. Washington, D. C.: Winston. [JStBTE, RAG]

Revlin, R., & Mayer, R. E., eds. (1978) *Human reasoning*. Washington, D. C.: Winston. [RAG, MH]

Revlis, R. (1975) Syllogistic reasoning: Logical decisions from a complex data base. In: *Reasoning: Representation and process*, ed. R. J. Falmagne, pp. 93–133. Hillsdale, N. J.: Erlbaum. [MH]

Ross, L. & Lepper, M. R. (forthcoming) The perseverance of beliefs: Empirical and normative considerations. In: *New directions for methodology of behavioral science: Fallible judgment in behavioral research*, ed. R. A. Shweder & D. Fiske. San Francisco: Jossey-Bass. [LJCa]

Russell, B. (1919) *Introduction to mathematical philosophy*. London: Allen & Unwin. [LJCa]

Sampson, G. R. (1975) *The form of language*. London: Weidenfeld & Nicolson. [LJCa, GS]

(1979) *Liberty and language*. Oxford: Oxford University Press. [GS]

Seligman, M. E. P. (1970) On the generality of the laws of learning. *Psychological Review* 77:406–18. [HJE]

Shackle, G. L. S. (1949) *Expectation in economics*. Cambridge: Cambridge University Press. [LJCr]

Shafer, G. (1976) *A mathematical theory of evidence*: Princeton: Princeton University Press. [DHK]

(1978) Non-additive probabilities in the work of Bernoulli and Lambert. *Archive for History of Exact Sciences* 19:309–70. [LJCr]

Shanteau, J. C. (1974) Component processes in risky decision judgments. *Journal of Experimental Psychology* 103:680–91. [LLL]

Shuford, E. H. (1959) A comparison of subjective probabilities for elementary and compound events. University of North Carolina Psychometric Laboratory Report #20. [LLL]

Simon, H. A. (1957) *Models of man*. New York: John Wiley. [LJCa]

(1978) Rationality as process and as product of thought. *American Economic Review* 68:1–16. [HJE]

Skyrms, B. (1980) *Causal necessity*. New Haven: Yale University Press. [BS]

Slovic, P.; Fischhoff, B. & Lichtenstein, S. (1976) Cognitive processes and societal risk taking. In: *Cognition and social behavior*, ed. J. S. Carroll & J. W. Payne. Hillsdale, N. J.: Erlbaum. [LJCa]

(1977) Behavioral decision theory. *Annual Review of Psychology* 28:1–39. [LJCa]

Smedslund, J. (1970) Circular relation between understanding and logic. *Scandinavian Journal of Psychology* 11:217–19. [JS]

(1977) Piaget's psychology in practice. *British Journal of Educational Psychology* 4:71–6. [JS]

(1978) Bandura's theory of self-efficacy: A set of common sense theorems. *Scandinavian Journal of Psychology* 19:1–14. [JS]

(in press a) The logic of psychological treatment. *Scandavian Journal of Psychology* 22: [JS]

(in press b) What is necessarily true in psychology? *International Journal of Theoretical Psychology: Annals* 1: [JS]

Sober, E. (1978) Psychologism. *Journal of the Theory of Social Behavior* 8:165–91. [WCL]

(1981) The evolution of rationality. *Synthese* 46:95–120. [WGL]

Sommers, F. T. (1981) *The logic of natural language*. Oxford: Oxford University Press. [LJCa]

Spinoza, B. de. (1914) *Opera*. Hague: Nijhoff. [LJCa]

Stich, S. P. (1975) Logical form and natural language. *Philosophical Studies* 28:397–418. [LJCa]

Stich, S. & Nisbett, R. (1980) Justification and the psychology of human reasoning. *Philosophy of Science* 47:188–202. [REN, SPS]

Swinburne, R. G. (1964) Falsifiability of scientific theories. *Mind* 73:434–36. [LJCa]

Toda, M. (1962) The design of a fungus-eater: A model of human behavior in an unsophisticated environment. *Behavioral Science* 7:164–83. [HJE]

Todhunter, I. (1949) *A history of the mathematical theory of probability from the time of Pascal to that of Laplace*. Reprint of 1865 edition. New York: Chelsea. [LJCar, PD, IN]

Tribe, L. H. (1973) Technology assessment and the fourth discontinuity: The limits of instrumental rationality. *Southern California Law Review* 46:617–60. [HJE]

Tversky, A. (1967) Utility theory and additivity analysis of risky choices. *Journal of Experimental Psychology* 75:27–36. [LLL]

Tversky, A. & Kahneman, D. (1971) The belief in the "law of small numbers." *Psychological Bulletin* 76:105–10. [LJCa]

(1973) Availability: A heuristic for judging frequency and probability. *Cog-*

nitive Psychology 5:207–32. [LJCa, RWW]
(1974) Judgment under uncertainty: Heuristics and biases. *Science* 125:1124–31. [LJCa, PD, RJS]
(1977) Causal thinking in judgment under uncertainty. In: *Basic problems in methodology and linguistics*, ed. R. Butts & J. Hintikka. Dordrecht: Reidel.
[LJCa, ILAM, IN]
(1980) Causal schemas in judgments under uncertainty. In: *Progress in Social Psychology*, ed. M. Fishbein, 1:49–72. Hillsdale, N. J.: Erlbaum. [SLZ]
(1981) The framing of decisions and the psychology of choice. *Science* 211:453–58. [AT]
Van Duyne, P. C. (1974) Realism and linguistic complexity. *British Journal of Psychology* 65:59–67. [LJCa]
(1976) Necessity and contingency in reasoning. *Acta Psychologica* 40:85–101. [LJCa]
Venn, J. (1888) *The logic of chance*. 3rd ed. London: Macmillan and Co. [IN]
Wagenaar, W. A. (1972) Generation of random sequences by human subjects: A critical survey of literature. *Psychological Bulletin* 77:65–72. [LJCa]
Wason, P. C. (1960) On the failure to eliminate hypothesis in a conceptual task. *Quarterly Journal of Experimental Psychology* 12:129–40. [LJCa, JStBTE, NEW]
(1966) Reasoning. In: *New horizons in psychology*, vol 1, ed. B. Foss. Harmondsworth: Penguin. [LJCa, JStBTE, RWW]
(1968) Reasoning about a rule. *Quarterly Journal of Experimental Psychology* 20:273–81. [LJCa]
(1969) Regression in reasoning? *British Journal of Psychology* 60:471–80. [PW]
(1977) Self-contradictions. In: *Thinking: Readings in cognitive science*, ed. P. N. Johnson-Laird & P. C. Wason, pp. 114–28. Cambridge: Cambridge University Press. [MH]
Wason, P. C. & Golding, E. (1974) The language of inconsistency. *British Journal of Psychology* 65:537–46. [PW]
Wason, P. C. & Johnson-Laird, P. N. (1970) A conflict between selecting and evaluating information in an inferential task. *British Journal of Psychology* 61:509–15. [PW]
(1972) *Psychology of reasoning: Structure and content*. London: Batsford. [LJCa, RJS, PW, NEW]
Wason, P. C. & Shapiro, D. (1971) Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology* 23:63–71. [LJCa]
Wertheimer, M. (1959) *Productive thinking*. Enl. ed. New York: Harper. [MH]
Wetherick, N. E. (1970) On the representativeness of some experiments in cognition. *Bulletin of the British Psychological Society* 23:213–14. [NEW]
(1971) "Representativeness" in a reasoning problem: A reply to Shapiro. *Bulletin of the British Psychological Society* 24:213–14. [NEW]
(1973) Review of Wason, P. C. and Johnson-Laird, P. N., "Psychology of reasoning, structure and content." (1972) *Bulletin of the British Psychological Society* 26:45–46. [NEW]
Woodworth, R. S. & Sells, S. B. (1935) An atmosphere effect in syllogistic reasoning. *Journal of Experimental Psychology* 18:451–60. [RAG]