RATIONALITY AND CHARITY*

PAUL THAGARD
Department of Humanities
University of Michigan, Dearborn

RICHARD E. NISBETT†
Department of Psychology
University of Michigan, Ann Arbor

Quine and others have recommended principles of charity which discourage judgments of irrationality. Such principles have been proposed to govern translation, psychology, and economics. After comparing principles of charity of different degrees of severity, we argue that the stronger principles are likely to block understanding of human behavior and impede progress toward improving it. We support a moderate principle of charity which leaves room for empirically justified judgments of irrationality.

Introduction. Quine (1960, pp. 59, 69; 1969, p. 46) has recommended a "principle of charity", according to which translation should preserve logical laws: we should translate a speaker's utterances in such a way as to avoid construing those utterances as contradictory or absurd. Recent methodological discussions of psychology have drawn on similar principles and proposed that we should not give an account of people's cognitive behavior which labels them as illogical or irrational (Sober 1978; Dennett 1978; Cohen 1979, 1981). Similarly, over the past thirty years, the discipline of economics and the interdisciplinary field of decision science have developed a strong tradition of interpreting individual and institutional choice behavior in such a way as to preserve the assumption of rationality (March 1978).

In this paper we shall assess the adequacy of principles of charity as canons of social science methodology. Our first task is to state more explicitly than is usually done what the various principles of charity are intended to enjoin. We display five degrees of severity of principles of charity, ranging from mild recommendations to be cautious in finding people to be irrational, to universal proscriptions against imputing irrationality. We shall argue that the stronger principles of charity are likely to block understanding of human behavior and impede progress toward

*Received June 1982; revised September 1982.
†For helpful comments, we are grateful to Denis Dutton, Allan Gibbard, Alvin Goldman, Daniel Hausman, Daniel Kahneman, Susan Kus, John McCumber, Daniel Moerman, and Stephen Stich. We also thank Ed Sammons for editorial assistance.

Philosophy of Science, 50 (1983) pp. 250–267.

improving it. In psychology, economics, and other fields, we may well have good reasons for interpreting the behavior of subjects as irrational. To show this, we shall first examine Quine's principle and argue on anthropological and other grounds that it may be legitimate to translate someone's utterances in such a way that they are contradictory or absurd. We then criticize attempts by Sober, Dennett, Cohen, and Davidson to import a strong principle of charity into psychology. We next criticize the strong presumption of rationality held by many in the fields of economics and decision theory. Then we defuse a possible political motivation for a principle of charity, arguing that it is not the innocent egalitarian principle that it appears to be. Finally, we support a moderate principle of charity, which leaves room for empirically justified judgments of irrationality.

I

It is necessary first to state the problem more clearly, by carefully formulating the relevant notions of rationality and the content of principles of charity. The term "rationality" has varied uses, many orthogonal to our concerns.[1] We are interested primarily in the sense of "rationality" in which a thought or action is rational if it conforms to the best available normative standards. Thus we define rational behavior as what people *should* do given an optimal set of inferential rules.

We can now state the most general principle of charity: Avoid interpreting people as violating normative standards. This formulation is much too vague, in two respects. First, we need to specify what normative standards are involved; specification will give rise to different principles of charity for translation, inferential behavior, and choice behavior. Here are principles of charity for each domain.

Translation: Avoid translating subjects' utterances in such a way as to accuse them of holding contradictory or absurd beliefs.

Inference: Avoid interpreting subjects' inferences in such a way as to accuse them of violating normative rules of deductive or inductive reasoning.

Choice: Avoid interpreting subjects' economic behavior or other choices as violating normative standards of decision making.

The second and more serious sort of vagueness in our initial principle

[1]For notions of rationality very different from the one we discuss, see Bennett (1964), Kekes (1976). Agassi and Jarvie have usefully distinguished three levels of rationality. At the lowest level, an agent's action is said to be rational if it is goal directed. At the next level, we call thinking rational if it obeys some set of explicit rules. At the third and most important level, a rational thought or action is one which conforms to the highest standards. We are interested in this last level. See Jarvie and Agassi (1970), Agassi and Jarvie (1979).

of charity concerns how stringently we are to take the injunction to "avoid" imputing irrationality. At least five levels of stringency can be distinguished, generating the following principles which range from the weak to the very severe:

(1) Do not assume a priori that people are irrational.
(2) Do not give any special prior favor to the interpretation that people are irrational.
(3) Do not judge people to be irrational unless you have an empirically justified account of what they are doing when they violate normative standards.
(4) Interpret people as irrational only given overwhelming evidence.
(5) Never interpret people as irrational.

Combining these five levels of stringency with the three domains of translation, inference, and choice, we arrive at fifteen principles of charity. Which are acceptable canons of methodology?

In general, we want to argue that adoption of principles of charity at levels (4) and (5) is not desirable. We shall challenge principles at these levels by displaying cases where the imputation of irrationality seems well justified on partially empirical grounds, and by undercutting the methodological and political rationales for principles of charity. Judgments of irrationality presuppose a set of normative principles, but we do not assume that these principles are true a priori, or otherwise immutable. Rational principles evolve and improve, like scientific theories. To judge behavior irrational is to say that the behavior violates the principles that we currently, objectively, hold. We shall argue that there is good reason to believe that people's behavior often violates normative principles so defined. Our judgments of irrationality are not made lightly: we espouse level (3) principles of rationality and accordingly impute irrationality only when there is an empirically plausible account of why people are not following normative standards.

II

Let us first consider principles of charity as they concern translation. In *Word and Object,* Quine makes the following comments: "fair translation preserves logical laws"; "assertions startlingly false on the face of them are likely to turn on hidden differences of language"; "one's interlocutor's silliness, beyond a certain point, is less likely than bad translation" (Quine 1960, p. 59). Quine seems to base these injunctions on the existence of "semantic criteria" for truth functions, criteria involving observable behavior of assenting to or dissenting from particular utterances. Negation turns a short sentence to which one will assent into a

sentence from which one will dissent, and vice versa, while conjunction produces compounds to which one is prepared to assent when one is prepared to assent to each component (Quine 1960, p. 57). If we had such behavioral means to identify the foreign analogues of our "not" and "and", then it is indeed hard to imagine a subject assenting both to a sentence and its negation. We would, as Quine suggests, be inclined to retranslate or come up with some other explanation of the subject's inclination to self-contradiction.

Quine's assertion that bad translation is only more "likely" than absurdity in those interpreted suggests that Quine might be using a moderate level (3) principle. However, his attempt to fix logical terms behaviorally suggests that he would recommend that we should never, or almost never, find people in violation of the laws of logic. He says that the claim that certain natives accept as true sentences that are translatable in the form '*p* and not *p*' is "absurd under our semantic criteria" (Quine 1960, p. 58). This quotation suggests that Quine would urge a level (5) principle of charity, according to which we should never translate utterances as contradictory. We shall now argue that adoption of a level (5), or level (4), principle of charity would be illegitimate in the case of translation.

Quine assumes that the dispositions to assent and dissent to sentences, on which our translations are based, are universal. It is quite conceivable, however, that in very many cases a subject will be inclined to assent and dissent in such a way as to establish very plausible translations of "not" and "and". Given these cases, what should we do when faced with a more or less isolated case where subjects are inclined to assent to both a sentence and its negation? Quine would have us abandon our well established translation, but, given the linguistic evidence we already have, it is equally plausible to decide that, in the matter at hand, the subject simply is inconsistent and is fully prepared to violate our laws of logic.

John Kekes describes a plausible example of violation of logical laws, based on Evans-Pritchard's discussion of Nuer religion (Kekes 1976). Kekes claims that the Nuer violate the principle of identity, since they affirm that swamp light is spirit, while denying that spirit is swamp light, where "is" can be seen to signify identity, not predication. Nevertheless, we can translate the inconsistent beliefs into English because we have a broad understanding of their language, belief system, and customs. As Kekes summarizes, "we can understand the illogical beliefs of Nuer because prediction, explanation, translation and imaginative emotional application of them are possible" (Kekes 1976, p. 383). Understanding illogical beliefs requires a much larger battery of hermeneutic techniques than Quine's behavioristic reliance on criteria of assent and dissent.

Ernest Gellner makes the strong claim that "the overcharitable interpreter, determined to defend the concepts he is investigating from the

charge of logical incoherence, is bound to misdescribe the social situation" (Gellner 1973, p. 39). He defends the claim by presenting a number of examples of cultures in which there appear to be conceptual contradictions, but where these contradictions play important social roles. For instance, he describes the status of *agurram* in the culture of central Moroccan Berbers, and argues that it is essential to the concept both that the holder of that status be accredited with certain characteristics, and that he should not really possess them. Whereas it is the general belief that people having *agurram* status are selected by God, it is socially important that they are in fact selected by the surrounding ordinary tribesmen who use their services. Despite its logical incoherence, the concept of *agurram* survives because it plays a significant role in dictating social behavior. We would misconstrue the situation if we neglected this social role and insisted on logical charity. Because language can have functions other than communication of truths, translation cannot always be charitable.

To assume that the Nuer, Berbers and other peoples are disinclined to assent to contradictions in all cases is to assume that they share our attitudes toward formal logic, but why expect this any more than we would expect them to accept our attitudes toward empirical science? We can well imagine a subject believing that formal logic applies in some domains but not in others, just as some scientists are devoutly religious. Quine seems to assume that to violate the laws of logic in some cases is to be prepared to violate them in *all* cases, for he observes that classical laws of logic yield all sentences as consequences of any contradiction (Quine 1960, p. 59). But nothing compels a subject—or us, for that matter—to believe all the logical consequences of our beliefs. Laws of logic tell us what we *may* infer, not what we *must* infer (Thagard 1979). Moreover, there are psychological reasons to suppose that our beliefs are organized into sub-systems in such a way that an inconsistency could be maintained in one sub-system without infecting other sub-systems.[2] Similarly, inconsistency could exist between sub-systems without necessarily introducing inconsistency within sub-systems, or at one level of generality without introducing inconsistency at other levels of generality. Hence one can intelligibly embrace a contradiction in one aspect of one's thinking, while operating in accord with logic in other aspects. The argument against level (4) and (5) principles of charity is that we can gain such a thorough knowledge of a subject's language through study of those other aspects that, when faced with an apparently contradictory utterance, we should construe it as contradictory rather than revise our well-established translations.

[2]P. Thagard, "Frames, Knowledge, and Inference", unpublished ms.; the argument originates with Marvin Minsky.

Anthropological examples are always open to the charge that we have insufficiently understood an alien people, and that further study would show that they do not really violate logical principles. Such question-begging claims are unanswerable, so let us look at examples of violations of logic closer to our own culture. A. V. Miller translates a sentence from Hegel's *Logic* as follows: "Something moves, not because at one moment it is here and at another moment there, but because at one and the same moment it is here and not here, because in this 'here', it at once is and is not" (Hegel 1969, p. 440). The context makes it clear that Hegel is intentionally challenging the principle of contradiction. He thinks that Zeno's paradoxes provide evidence of the contradictory nature of motion. To understand Hegel's assertion, we need a full account of his view of dialectics, particularly of his complex notions of dialectical negation and contradiction (Thagard 1982c). Indeed, Hegel's notions of negation and contradiction are different from Quine's truth functional ones, but, in the sentence quoted, Hegel is using the German equivalents of "and" and "not" in ways so familiar that no other translation would be appropriate. We have to understand and translate Hegel as violating the principle of contradiction. To do so is not to be "uncharitable", but merely to take him seriously as a complex and iconoclastic thinker.

Other examples from the history of philosophy could be given to show the dangers of interpreting philosophers on the basis of our current logical principles. Mystical philosophies such as Zen are particularly likely to be misunderstood if a principle of charity is applied. Why should we withhold judgments of absurdity from mystics who profess to revel in it?

Behind the translational principle of charity is the assumption that the primary function of language is usually communication, so that people are using language to convey information. But in many contexts, from Zen to cultural rituals, language can be used for social or other ends. Maintenance of principles of logic may be irrelevant to such ends.

Hence translation of the utterances of people of other cultures and philosophies may well *require* that we understand them as violating our principles of logic, of rationality. Strong translational principles of charity are therefore empirically unsound. We shall now argue that similar principles of charity are also inappropriate for understanding the ordinary inferential behavior of people in our own culture.

III

Contemporary cognitive psychologists study human thinking as a kind of information processing. They ask: What processing mechanisms explain our cognitive performance? It might be presumed that one essential mechanism would be an inferential system that serves to derive new be-

liefs from old ones in accord with principles of logic. If people had such a mechanism, then strong inferential principles of charity, at level (4) or (5) might be in order; we should not give accounts of human information processing which suppose that people violate logical rules.

But why suppose that people possess such a mechanism? One might assume that natural selection has endowed us with a built-in inferential mechanism operating in accord with the laws of logic.[3] But it is not at all evident that such an elaborate mechanism was necessary for survival in the environments in which human beings and their predecessors evolved. How much logic do you need to hunt and gather? You need some, but not enough to guarantee that you will shrink from violating the principle of contradiction, and certainly not enough that you will be able to manage complex hypothetical and statistical reasoning. We can suppose that natural selection will favor some sort of inferential system which incorporates much of logic, but there is no reason to suppose that *all* our logical principles are hard-wired (Sober 1980).

The argument that *inductive* inferential principles are hard-wired is particularly difficult to sustain. Hacking has argued that the modern conceptions of probability and evidence are scarcely 300 years old (Hacking 1975). These conceptions, in particular statistical principles such as the law of large numbers and the principle of regression, are essential for reasoning adequately about a host of problems. The work of Tversky and Kahneman (1974) and others shows that people regularly violate probabilistic and statistical rules in their reasoning not just about scientific matters but in their reasoning about everyday affairs as well.

Nisbett and Ross (1980) reviewed the large body of evidence on lay inductive reasoning and identified a number of very pervasive errors. We will mention just three of these. Kahneman and Tversky (1972) and many subsequent investigators have shown that people frequently fail to recognize that large deviations from population parameters are more likely for small samples than for large samples. For example, subjects report believing that atypical proportions of male births are no more likely at hospitals with an average of 15 births per day than at hospitals with an average of 45 births per day. Similarly, people often do not recognize that, under conditions of imperfect predictability from one variable to another, extreme values on one variable will be associated with values on the other variable that are less extreme (Kahneman and Tversky 1973). For example, subjects predict that a target person's grade point average is extremely high both when given the information that the target has very

[3]This assumption appears to be made by several writers, e.g. Dennett (1978, 1981, forthcoming) and Lycan (1981). It has been effectively criticized by Stich (forthcoming) and by Einhorn and Hogarth (1981).

high intelligence test scores and when given the information that the target has a very good sense of humor. An extreme prediction is justified only when correlations are believed to be very high, which is not the case for subjects' beliefs about the association between sense of humor and grade point average. Another general error is people's frequent inability to recognize or compensate for statistical bias in evidence. For example, Hamill, Wilson and Nisbett (1980) showed that subjects' beliefs about welfare recipients became more negative after reading about a particularly squalid welfare case, and did so to the same extent whether the case was described as typical of recipients generally, with respect to length and degree of dependence on welfare, or as highly atypical.

The evidence indicates that people make inferential errors. The errors seem to be due to lack of knowledge of certain inductive rules or an inability to apply them. If so, then people are not fully rational in that their inferences fall short of the best available normative standards.

Nevertheless, several philosophers have argued, on different methodological grounds, that an assumption of rationality is required for the explanation of human inferential behavior. Dennett claims that *any* intentional system must be supposed to follow the rules of logic.[4] To adopt the intentional stance toward a system is to assume that it is rational, that is, that it has an "optimal design relative to a goal" (Dennett 1978, p. 5). Dennett claims that intentional prediction is only possible if we assume that the system is capable of using logic to draw out the consequences of its beliefs and thereby act in accord with our expectations.

There is no reason, however, why it should not be possible to determine empirically that a system is regularly using some inferential principle or heuristic that departs from standard logical principles, then to use the operation of this heuristic as part of an explanation of the system's behavior. This is the strategy that seems to be required in light of the recent empirical work on inductive reasoning. The work provides substantial evidence that people employ simple inferential heuristics that are useful for many problems but inadequate for others requiring more complicated inductive rules or statistical principles. Errors result. As Stich has said of the proper interpretation of these errors: "In using intentional locutions we are presupposing that the person or system to which they are applied is, in relevant ways, similar to ourselves. Thus inferential errors that we can imagine ourselves making—errors like those [uncovered by Kahneman and Tversky and Wason and Johnson-Laird (1972)] —can be described comfortably in intentional terms. It is only the sort of error or incoherence that we cannot imagine falling into ourselves that

[4]D. Dennett (1978), p. 11. Of people on another planet, Dennett insists (p. 9): "in virtue of their rationality they can be supposed to share our belief in logical truths."

undermines intentional description" (Stich forthcoming, p. 9). An assumption of rationality might be a useful first approximation before we have enough evidence to construct a full account of the system's cognitive behavior, but we should be prepared to construct an alternative account if the behavior of the system seems less than rational, in ways that are familiar.

Sober has recognized that whether to construe human cognitive behavior as rational is an empirical question (Sober 1978), and he proposes a slightly different reason for favoring rationality. He says that, when faced with apparently irrational behavior, we can postulate two sorts of models. In one, we assume humans to be inherently rational, with deviations from rationality explained by various forms of interference caused by adverse conditions. In the other, we postulate a fundamental mechanism which is not perfectly rational. Sober claims that the first sort of model is to be preferred on grounds of *simplicity*. One gets a simpler view of cognition by using valid rules that are subject to interference, since the overwhelming majority of inferences that people make are valid.

But is that latter claim true? The recent psychological evidence suggests that people frequently make invalid inferences of certain identifiable types. In a domain in which people can be shown to deviate regularly from normative principles, Sober's claim that it is simpler to assume rationality would no longer hold. As we saw with translation, the evidence may mount sufficiently to require us to drop the rationality assumption. If we hold onto level (4) or (5) principles of charity, we may find ourselves postulating ad hoc interferences to maintain the presumption of rationality, and the ensuing epicycles may well render the rationality model more complicated than a model which sees a less than logical mechanism operating.

Sober bases much of his case for the simplicity of the rationality assumption on an analogy with Chomskian linguistics (Sober 1978). Native speakers of a language can be assumed to have grammatical *competence,* even if occasionally their *performance* deviates from grammatical standards. Similarly, Sober argues, it is simpler to explain away deviations from inferential standards by assuming that interference and error have made the subject's performance belie his or her inherent competence.

Jonathan Cohen also invokes a competence/performance distinction in support of a principle of charity. He asserts:

> . . . 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 human beings a

> cognitive competence—however often faulted in performance—that corresponds point by point with the normative theory (Cohen 1981, p. 321).

Cohen supposes that people's intuitions about what is inferentially valid are on a par with native speakers' intuitions about what is grammatically correct. Then, just as we assume that native speakers are competent linguistically, we must assume that reasoners are competent logically, even if their performance is not always optimal.

Whatever the value of the competence/performance distinction in linguistics, it is not appropriate for questions of logic and rationality. We can legitimately assume that every speaker of a language tacitly knows the grammar of that language; that assumption is virtually tautologous. But on what basis do we assume that people in general have competence in reasoning? We have no reason to believe that inferential competence is innate, and we have no reason to suppose that, unlike linguistic knowledge, it is acquired by all members of a culture. Indeed, whereas we might suppose people in a given linguistic community to be roughly on a par in their ability to recognize and utter grammatical sentences (although even here there are real differences [Stich forthcoming]), we find a wide range of abilities to perform deductive and inductive inferences. The discrepancy is less evident in the case of simple deductive inference, where almost everyone can handle modus ponens, but is dramatic in the case of inductive reasoning. Among university students without formal training in statistics or probability theory there are marked individual differences in ability to reason about everyday problems in accordance with statistical principles (Jepson, Krantz and Nisbett, forthcoming). Moreover, people's ability to use a statistical approach for both scientific and everyday problems improves significantly with training (Fong, Krantz and Nisbett, forthcoming). Any assumption of a general inferential competence, even among educated adults in our own culture, is therefore lacking in empirical foundation.

An equally important argument against the analogy to linguistics is the fact that improvement in standards for inductive reasoning is possible. We cannot say that English grammar has improved over the past 400 years, but inferential techniques clearly have. Modern symbolic logic provides methods for assessing the validity of complex arguments which are much more powerful than Aristotelian methods, which were largely restricted to the syllogism. Inductive advances are even more striking: the gains of the seventeenth century were not the result of a magical improvement in people's intuitions, but of the development of sophisticated and effective models of reasoning against which our intuitions can be

evaluated. The body of inductive rules employed by statisticians and scientists, and the domains to which they can be applied, are undergoing constant growth and change.

Rationality is therefore not a static concept: rational inference is inference in accord with the best available rules, and our set of rules is constantly being improved. The fact that there is progressive development of rational rules demonstrates that Cohen errs in taking general inferential intuitions (whether this is construed as a mythical universal competence or a momentary majority competence) as ultimate data for normative theory. Intuitions play a role in the development of normative theory, but the role is highly complex and dynamic: intuitions may suggest what is normatively correct, but we can use innovations in normative theory to override and improve our intuitions (Goodman 1965, Stich and Nisbett 1980, Thagard 1982b). We are then in a position to criticize as less than rational the judgments of those whose inferential practice has not similarly improved. Since competence in reasoning to the current high standards of mathematical logic and statistical inference needs to be taught, we can attach no special primacy to people's untutored intuitions about validity, nor can we grant that it is simpler to assume that people are rational in their inferences.

To conclude this section, we must mention an astoundingly strong kind of charitable principle advocated by Donald Davidson and Daniel Dennett. They both urge that understanding of others requires not only that we assume their inferential practices to be rational, but also that we assume their beliefs to be *true*. Davidson asserts: "Charity is forced on us—whether we like it or not, if we want to understand others, we must count them right in most matters" (Davidson 1973/74, p. 19). Dennett holds that the beliefs of an intentional system must be presumed to be "those it ought to have, given its perceptual capacities, its epistemic needs, and its biography" (Dennett 1981, p. 42). Davidson is particularly concerned to uphold the assumption of correctness with respect to understanding people's beliefs about mental events:

> People are in general right about the mental causes of their emotions, intentions, and actions because as interpreters we interpret them so as to make them so. We must, if we are to interpret them at all (Davidson 1976, p. 757).

The inadequacy of the philosophical underpinning of Davidson's theses has already been discussed by Colin McGinn (1977), and we shall not repeat his arguments here.[5] What we wish to note is the empirical in-

[5]According to McGinn, Davidson's main basis for espousing a principle of charity is that without assuming that most of a subject's beliefs were true, we would not be able to

adequacy of Davidson's and Dennett's principles, anthropologically and psychologically. Cultural anthropology does indeed require that we approach radically different belief systems with as great as possible a suspension of our own presuppositions, but nothing in the hermeneutic process requires us actually to accept the presuppositions of the exotic culture under study. We can understand a people's belief that swamp light is spirit, or that sex and procreation are unrelated, without supposing that their belief systems have any truth at all. Davidson's assertion about mental causes is also a straightforward empirical one, and recent evidence indicates that the assertion is false. It is possible to investigate the real causes of people's emotions and actions by empirical means, and then to question people about their beliefs about the causes of their emotions and actions. Work employing this strategy shows that people are often wrong. They assert both that influential factors were not influential and that non-influential factors were influential (Schachter and Singer 1962, Nisbett and Wilson 1977). Work in social science supports the view that people's beliefs are often empirically wrong even more clearly than it does the view that people are often irrational.

IV

There is also a strong tradition of assuming rationality in economic decisions and other choices. As March puts it:

> Much theoretical work searches for the intelligence in apparently anomalous human behavior. This process of discovering sense in human behavior is conservative with respect to the concept of rational man and to behavioral change. It preserves the axiom of rationality; and it preserves the idea that human behavior is intelligent, even when it is not obviously so (March 1978, p. 589).

Economics and decision science rest on a strongly-held normative model of rational choice, that of expected utility theory.[6] This model has been widely applied as a descriptive model of actual economic behavior (Friedman and Savage 1948, Arrow 1971). Some theorists, since Simon's classic work (1957), have dropped the presumption that the normative theory generates good predictions for all types of actual behavior, but have explained departures from normative standards in ways that preserve inherent rationality. The departures are seen as due to intrinsic limits, for

---

begin to figure out what beliefs the subject had. McGinn points out that our observation that subjects causally interact with objects in the environment gives us an independent start on the ascription of belief, obviating Davidson's excessive charity.

[6]We note, however, that philosophers such as Jeffrey (1977), Lewis (1981), and Gibbard and Harper (1978) have proposed more complex decision theories.

example, on human computational capacities, or as due to external constraints on ability to maximize expected utility. March (1978) listed seven distinct types of constraint assumptions that have been made to preserve the principle of rationality for general human choice strategies. These assumptions make it possible to account on an ad hoc basis for almost any conceivable departure from normative standards and to leave intact the fundamental presumptions of expected utility theory.

We acknowledge that the case for charity is in general stronger for choice than for inductive inference. One can gain some confidence in assessing the validity of a given induction by analyzing whether it seems to employ legitimate rules and reliable information. But in the case of choice, very many beliefs, as well as many desires, may underlie every action, and the possibility is great that some background belief or desire might be found to justify an apparently irrational choice. Moreover, as March suggests, a variety of constraints, including unconscious ones, limit our ability to make full use of the apparatus of decision theory. Since it is virtually impossible to rule out all these factors in any given case of apparent irrationality, the presumption of rationality is extraordinarily hard to overturn.

Nevertheless, there are good reasons for doubting whether a strong level (4) or (5) principle of charity is appropriate even in the case of choice. First, "charitable" interpretations of behavior can be forced and implausible. For example, critics of experiments on betting behavior often defend the suboptimal behavior of subjects by asserting that the subjects are maximizing not monetary gain but self-esteem or amusement. Such alternative explanations may be correct, but we should not be obliged to entertain them regardless of independent empirical corroboration. A second reason to limit a choice principle of charity is that it provides more prior buttressing to current versions of decision theory, in both its normative and descriptive aspects, than any theory ought to be allowed. In fact, Kahneman and Tversky (1979, 1981) have presented evidence indicating that normative decision theory is not even a good first approximation to a descriptive model. They argue that people assign value primarily to gains and losses from their current reference point rather than to final assets and that values are not multiplied by probabilities at all but rather by "decision weights" that do not obey the probability axioms. Their work thus indicates that people commonly diverge from the current normative standards of decision making. Perhaps this shows that in some respects utility theory is in need of revision or replacement as the normative standard (Thagard 1982a). In any case, we seem to have in the case of decision making the same kind of systematic departure from normative principles which characterizes the sorts of inferences discussed in the last section. We should not be forced to paper over such possible

departures with empty charitable pronouncements. To do so would be to place a roadblock in the way of possible future discoveries concerning how people actually make choices.

A third reason for rejecting strong principles of charity for choice and action applies to inferences as well. A rampant principle of charity preempts the possibilities of criticism and improvement. If we cannot assume actions and judgments to be irrational, then we cannot hope to educate and improve choice strategies and inferential procedures. A heavy-handed charity principle would freeze human behavior in an unprogressive amalgam of late twentieth century procedures.

V

It's nice to be charitable. Readiness to impute irrationality smacks of elitism and ethnocentrism. Who are we to tell the common people that their inferential practices are inadequate? Who are we to tell an alien people that their cosmological beliefs are unfounded? Principles of charity are of course consonant with twentieth century philosophical and social scientific relativism. One contributor to this relativism has been a rising egalitarianism, which rejects the claims to hegemony of a European elite. This well-founded political egalitarianism does not conflict with our rejection of strong principles of charity: political egalitarianism is consistent with cognitive elitism.

Why suppose otherwise? We trace the error to right-wing arguments that attempt to justify political inequality on the basis of intellectual inequality, rooted in racial or individual differences. It is natural to challenge such arguments by replying first that there is little evidence for the alleged intellectual differences, and then implying that because people *are* intellectually equal, at least potentially, they should have equal say in the democratic process. However, the second part of the reply is dangerously wrong, since it accepts the assumption of the right wing argument that political equality is somehow a function of intellectual equality. The democratic principle that people universally should have equal rights and opportunities in no way requires that each existing individual be fully rational, any more than it requires that they each be six feet tall or as strong as a decathlon champion.

Of course, rationality *is* desirable. The conclusion to be drawn from empirical findings that people frequently fail to meet our developed rational standards is not that they should be disqualified from political deliberations. The appropriate lesson is rather that society has an obligation to educate its members in inferential techniques. Some training in logic and statistics should be part of the background of every educated person, and it should be an ideal of a democratic society that every person be

educated. In order for an individual to play a fully responsible role in the operation of a democratic society, he or she needs to be able to assess critically the host of political and empirical claims made concerning policy issues. Education in logical and statistical techniques is part of what is needed to develop that sort of critical faculty.

The assumption that humans are fully rational is very old, and it is possible to argue that it is largely pernicious in its consequences. Dawes has noted that Western theories of personality, from Aristotle to Maslow, have been hierarchical, with the rational faculties at the top, inviolate and unerring in themselves, but subject to interference from the passions below (Dawes 1976). Such a theory of personality would appear to have the unfortunate result that it encourages us to assume that when others see things differently from the way we do, or take actions that injure ourselves or our fellows, they must do so for reasons of self-interest. We are then encouraged to attribute their behavior, not to intellectual failures, but to immorality. The presumption of immorality is a much more effective call to arms than the presumption of irrationality. Therefore the belief that people's cognitive capacities are fully rational may foster conflict, resentment, and distrust.

True charity may result from accepting the conclusion that people are not always fully rational. A beneficent humility may result from the assumption that this conclusion also applies to oneself.

VI

Having defused the political motivation for strong principles of charity, let us quickly review the chief arguments we gave against applying such principles in anthropology, psychology, and economics. The central claim was that whether people's behavior diverges from normative standards is an empirical question. Hence, severe strictures concerning what degree of rationality to expect run the risk of hampering the discovery of such divergences. We have presented examples from the various fields which suggest that discrepancies between normative standards and behavior are sufficiently great that strong principles of charity are indeed hindrances to understanding human behavior.

In addition to their scientific inapplicability, principles of charity may be socially undesirable. The assumption that people *must* be fundamentally rational blocks the possibility of systematic education to eliminate the gap between behavior and normative standards. On the other hand, recognition that people frequently are inferentially inadequate opens the door for educational programs which can substantially improve general inferential performance.

A third reason for resisting principles of charity is that their application

can hinder development of new normative principles. Discovery of discrepancies between inferential behavior and normative standards may in some cases signal a need for revision of the normative standards, and the descriptions of behavior may be directly relevant to what revisions are made (Goodman 1965, Stich and Nisbett 1980, Thagard 1982b; see further Goldman 1978). Hence on scientific, social, and logical grounds, we recommend the exclusion of strong principles of charity from social scientific methodology.

We are not, however, urging the total abandonment of principles of charity. Implicit in the examples we used to show that the imputation of irrationality is sometimes justified were third level principles of charity:

(3) Do not judge people to be irrational unless you have an empirically justified account of what they are doing when they violate normative standards.

Principles at levels (1) and (2), according to which we should not attach any prior favor to judgments of irrationality, are too weak to be helpful. Judgments of irrationality should not be made lightly. We recommend finding people in violation of normative standards only when an alternative account of their behavior can be established. Thus in our anthropological examples, we suggested that contradictory remarks are to be understood in terms of their social function. We translate Hegel as making a contradictory utterance because we know he has reasons for wanting to reject the principle of contradiction. Similarly, in the realm of inference, it is legitimate to find people falling short of normative standards if we can describe what procedures they are using in lieu of accepted normative ones. Since we can determine that subjects who lack statistical instruction tend to use more familiar deterministic sorts of reasoning, we can appropriately judge them to be in violation of the normative statistical standards. We were more tentative in finding decision makers irrational, since an account of what people are doing when they violate standard canons of decision making is still under development (Kahneman and Tversky 1979, 1981), and revision of currently accepted canons may be in order.

Thus moderate principles of charity do have a modest role to play in social science methodology. But we should be wary of attempts to use "charity" as a slogan to preempt judgments of irrationality.

REFERENCES

Agassi, J. and Jarvie, I. C. (1979), "The Rationality of Dogmatism", in T. Geraets (ed.), *Rationality Today*. Ottawa: University of Ottawa Press. pp. 353–362.

Arrow, K. J. (1971), *Essays in the Theory of Risk-Bearing*. Chicago: Markham.

Bennett, J. (1964), *Rationality*. London: Routledge and Kegan Paul.

Cohen, L. J. (1979), "On the Psychology of Prediction: Whose is the Fallacy?", *Cognition* 7: 385–407.

Cohen, L. J. (1981), "Can Human Irrationality be Experimentally Demonstrated?", *Behavioral and Brain Sciences 4*: 317–331.
Cooper, D. E. (1975), "Alternative Logic in 'Primitive Thought'", *Man 10*: 238–256.
Davidson, D. (1973/74), "On the Very Idea of a Conceptual Scheme", *Proceedings of the American Philosophical Association 47*: 5–20.
Davidson, D. (1976), "Hume's Cognitive Theory of Pride", *Journal of Philosophy 73*: 744–757.
Dawes, R. M. (1976), "Shallow Psychology", in J. S. Carroll and J. S. Payne (eds.), *Cognition and Social Behavior*. Hillsdale, N.J.: Erlbaum.
Dennett, D. (1978), *Brainstorms*. Montgomery, Vermont: Bradford Books.
Dennett, D. (1981), "Three Kinds of Intentional Psychology", in R. Healey (ed.), *Reduction, Time, and Reality*. Cambridge: Cambridge University Press, pp. 37–61.
Dennett, D. (forthcoming), "True Believers: The Intentional Strategy and Why it Works", forthcoming in a volume of Herbert Spencer Lectures, Oxford University Press.
Einhorn, H. and Hogarth, R. M. (1981), "Behavioral Decision Theory: Processes of Judgment and Choice", *Annual Review of Psychology 32*: 53–88.
Fong, G. T., Krantz, D. H. and Nisbett, R. E. (forthcoming), "Effects of Statistical Training on Inductive Reasoning".
Friedman, M. and Savage, L. J. (1948), "The Utility Analysis of Choices Involving Risks", *Journal of Political Economy 56*: 279–304.
Gellner, E. (1973), *Cause and Meaning in the Social Sciences*. London: Routledge and Kegan Paul.
Gibbard, A. and Harper, W. (1978), "Counterfactuals and Two Kinds of Expected Utility", in C. Hooker, J. Leach, and E. McClennan (eds.), *Foundations and Applications of Decision Theory*, vol. 1, Dordrecht: Reidel.
Goldman, A. I. (1978), "Epistemics: The Regulative Theory of Cognition", *Journal of Philosophy 75*: 509–523.
Goodman, N. (1965), *Fact, Fiction, and Forecast*, second edition. Indianapolis: Bobbs-Merrill.
Hacking, I. (1975), *The Emergence of Probability*. Cambridge: Cambridge University Press.
Hamill, R., Wilson, T. D., and Nisbett, R. E. (1980), "Insensitivity to Sample Bias: Generalizing from Atypical Cases", *Journal of Personality and Social Psychology 39*: 578–589.
Hegel, G. (1969), *Science of Logic*. Trans. by A. V. Miller. New York: Humanities Press.
Jarvie, I. C. and Agassi, J. (1970), "The Problem of the Rationality of Magic", in B. Wilson (ed.), *Rationality*. New York: Harper & Row, pp. 172–193.
Jeffrey, R. (1977), "Savage's Omelet", in F. Suppe and P. Asquith (eds.), *PSA 1976*, vol. 2. East Lansing: Philosophy of Science Association, pp. 361–371.
Jepson, C., Krantz, D. H. and Nisbett, R. E. (forthcoming), "Inductive Reasoning: Competence or Skill?"
Kahneman, D. and Tversky, A. (1972), "Subjective Probability: A Judgment of Representativeness", *Cognitive Psychology 3*: 430–454.
Kahneman, D. and Tversky, A. (1973), "On the Psychology of Prediction", *Psychological Review 80*: 237–251.
Kahneman, D. and Tversky, A. (1979), "Prospect Theory: An Analysis of Decision Under Risk", *Econometrica 47*: 263–291.
Kahneman, D. and Tversky, A. (1981), "The Framing of Decisions and the Psychology of Choice", *Science 211*: 453–458.
Kekes, J. (1976), *A Justification of Rationality*. Albany: State University of New York Press.
Lycan, W. G. (1981), "'Is' and 'Ought' in Cognitive Science", *Behavioral and Brain Science 4*: 344–345.
March, J. G. (1978), "Bounded Rationality, Ambiguity, and the Engineering of Choice", *Bell Journal of Economics 9*: 587–608.
McGinn (1977), "Charity, Interpretation, and Belief", *Journal of Philosophy 74*: 521–535.
Nisbett, R. and Ross, L. (1980), *Human Inference: Strategies and Shortcomings of Social Judgment*. Englewood Cliffs: Prentice Hall.

Nisbett, R. E. and Wilson, T. (1977), "Telling More Than We Can Know: Verbal Reports on Mental Processes", *Psychological Reveiw 84*: 231–259.
Quine, W. V. O. (1960), *Word and Object*. Cambridge: MIT Press.
Quine, W. V. O. (1969), *Ontological Relativity and Other Essays*. New York: Columbia University Press.
Schachter, S. and Singer, J. E. (1962), "Cognitive, Social, and Physiological Determinants of Emotional State", *Psychological Review 69*: 379–399.
Simon, H. A. (1957), *Models of Man*. New York: Wiley.
Sober, E. (1978), "Psychologism", *Journal for the Theory of Social Behavior 8*: 165–191.
Sober, E. (1980), "The Evolution of Rationality", paper read to Society for Philosophy and Psychology.
Stich, S. (forthcoming), "Could Man Be an Irrational Animal?"
Stich, S. and Nisbett, R. (1980), "Justification and the Psychology of Human Reasoning", *Philosophy of Science 47*: 188–202.
Thagard, P. (1979), "In Defense of Deductive Inference", *Australasian Journal of Philosophy 57*: 274–279.
Thagard, P. (1982a), "Beyond Utility Theory", in M. Bradie and K. Sayre (eds.), *Reason and Decision*. Bowling Green, Ohio: Bowling Green State University, pp. 42–49.
Thagard, P. (1982b), "From the Descriptive to the Normative in Psychology and Logic", *Philosophy of Science 49*: 24–42.
Thagard, P. (1982c), "Hegel, Science, and Set Theory", *Erkenntnis 18*: 397–410.
Thagard, P. (forthcoming) "Frames, Knowledge, and Inference", unpublished manuscript.
Tversky, A. and Kahneman, D. (1974), "Judgment Under Uncertainty: Heuristics and Biases", *Science 185*: 1124–1131.
Wason, P. C. and Johnson-Laird, P. N. (1972), *Psychology of Reasoning: Structure and Content*. London: Batsford.