2

Rationality, prediction, and autonomous choice

Principles of rationality are invoked for several purposes: they are often deployed in explanation and prediction; they are also used to set standards for rational health for deliberating agents or to furnish blueprints for rational automata; and they are intended as guides to perplexed decision makers seeking to regulate their own attitudes and conduct. These purposes are quite different. It is far from obvious that what serves well in one capacity will do so in another. Indeed, I shall argue later on in this essay that when principles of rationality are intended for use as norms for self-criticism, they cannot also serve as laws in explanation and prediction or as blueprints for rational automata.

Whether such principles are used for explanatory purposes, for setting standards or for self policing, issues arise as to the scope of applicability of principles of rationality. Such principles are often employed to evaluate decisions as well as the attitudes which allegedly inform them. Whether principles of rationality are relatively weak constraints of 'coherence' and 'consistency' on belief, desire, probability judgment and other propositional attitudes is by no means settled. Many writers argue to the contrary. They insist that norms of rationality should be 'thickened' to include more substantive specifications of the beliefs, values, probability judgments which agents should have. At the thicker end of the spectrum, for example, is found the idea that morality and politics should somehow be derived from a conception of rationality, that factual beliefs and probability judgments should be rationally determined by 'evidence.'[1]

The question of scope raises issues not only about how strong principles of rationality should be but also about the domain of applicability. Many sympathizers with the Humean perspective doubt that 'the passions' can or should be subjected to even thin constraints of rationality in the manner in which it is alleged that beliefs can be criticized with respect to consistency.

Thanks are due to Carole Rovane for shrewd and patient criticism on matters of substance and style.

From *Canadian Journal of Philosophy*, Supp. Vol. 19 (1989), pp. 339–63. Reprinted by permission of The University of Calgary Press.

In the face of the many possible positions which can be taken on these matters, proposing an account of *the* concept of rationality is either foolhardy or presumptuous. I prefer instead to explain my motives for focusing on one kind of understanding of rationality.

My interest in rationality (or whatever one may wish to call it) derives from a fascination with the Peirce-Dewey 'belief-doubt' model of inquiry. I shall first seek to explain a way of understanding principles of rationality suited to a systematic account of the core features of well conducted inquiry according to the belief-doubt model. This discussion will then serve as background for my main contention that principles of rationality designed for use in self criticism cannot simultaneously be used for explanatory and predictive purposes.

The belief-doubt model begins with a distinction between what is taken for granted (i.e., is certain or fully believed) and what is conjectural (i.e., possibly false or open to doubt). It contends that the inquiring agent is in no need to justify what he or she fully believes. Justification is required for *changing* one's state of full belief - i.e., modifying the distinction between certainty and conjecture. On the one hand, inquirers seek to remove doubt and thereby convert conjectures into certainties and, on the other hand, there are occasions where inquirers should come to doubt what they initially took for granted. In the absence of a good reason to change, the inquirer should retain the commitments he has.[1]

According to this vision, the central problem is to give an account of well conducted (i.e., rationally conducted) inquiries leading to modifications of the divide between what is settled and what is doubtful.

A central feature of the Peirce-Dewey view of such inquiry is the

[1] To avoid any misunderstanding, it should be emphasized that this claim does not imply that the inquiring agent should regard himself as justified in suppressing the views of those with whom he disagrees. We may tolerate the public expression of dissent while expressing our own contempt for it. But we are not obliged to register contempt either. Indeed, the agent might sometimes listen to the dissenter's view even though those views seem absurd. Teachers are often obliged professionally to pay attention to points of view which they judge to be patently false and, indeed, to pretend that their minds are open when they are not. Such insincerity may be justified if it is effectively employed to induce students to question their views. And where professional obligation does not support such dissembling, the demands of civility in discourse may. We should, however, distinguish between contemptuous toleration both when the contempt is overt and when it is disguised for pedagogical purposes or due to considerations of conversational etiquette and genuine respect for dissenting views. Respect for a dissenting view arises when an agent confronted with dissent recognizes a good reason for genuinely opening his or her mind by ceasing to be convinced of the view initially endorsed which is in conflict with the dissent. Liberals of the sort I admire tolerate dissent but their toleration is often contemptuous. To confuse toleration with respect for the views of dissenters can lead advocates of toleration to urge upon us the skepticism of the empty mind.

assumption that justifying changes of states of full belief (and, hence, of doubt) is a species of attempting to adjust means to ends. The inquiring agent seeks to answer an as yet unsettled question. The conclusion of such an inquiry when an answer is obtained is a 'judgment' in Dewey's idiosyncratic terminology. The conjectures which constitute potential solutions of the problem under scrutiny are 'propositions' and, hence, are to be regarded as entertainable *means* to the given *end* of solving the problem. These means include not only conjectures or potential solutions to the problem but also the settled background information, techniques and methods taken as noncontroversial resources in the context of the inquiry. According to Dewey, propositions are affirmed and judgments asserted. When the inquiry is properly conducted with the appropriate adaptation of means to ends, Dewey called the assertion or judgment which represents the solution to the problem 'warranted.' The judgment which is a warranted assertion of one inquiry may then become a resource for subsequent deliberation and inquiry. Qua means of the new inquiry, it is not a warranted assertion. It is an affirmed proposition. This does not mean that it becomes a conjecture (unless called into question in the course of inquiry) but merely that background information, like conjectures, constitute ingredients of the instrumentalities of ongoing inquiry.[2]

[2] John Dewey (1938, pp. 118–20, 1941, pp. 270–2).

According to Dewey, every case of knowledge 'is constituted as the outcome of some special inquiry. Hence, knowledge as an abstract term is a name for the product of competent inquiries' (Dewey, 1938, p. 8). In Levi (1983, pp. 27–8), I suggested that Dewey should have abandoned this conception of knowledge. He should not have insisted on equating knowledge with warranted assertibility if he also wished to acknowledge, as Peirce had, that to be settled, a conviction did not require a prior justification. It mây, perhaps, be thought that an appreciation of the context sensitivity of Dewey's conception of 'warranted assertibility' would mitigate the difficulty. But if a proposition used as background information in an inquiry is not a warranted assertion relative to some prior inquiry, it is not knowledge even in the current inquiry in Dewey's official sense. Since Dewey is quite clear that his sense of 'knowledge' is honorific, withholding the epithet is tacit attribution of some deficiency. But it is precisely the allegation of the existence of such a deficiency that Peirce sought to rebut in 'The Fixation of Belief.' Knowledge and warranted assertibility cannot coincide as Dewey suggests. There can be propositions used as background information and, hence, qualifying as knowledge which are not products of prior properly conducted inquiries and, hence, are not warranted assertions relative to any historically conducted inquiries. Moreover, it can even happen that knowledge fails even for items that are warrantedly assertible relative to one inquiry provided the result claimed has been legitimately called into question in subsequent inquiry.

It may, perhaps, be worth mentioning that Dewey tended to think of conjectures or hypotheses as what I have tended to call 'potential answers' to the question under investigation. For me the distinction between settled assumptions and what is open to doubt or conjectural is presupposed by the deliberating or inquiring agent when facing a problem. It is one of the tasks of inquiry addressing a certain question, to identify

I do not suggest that we adopt Dewey's nonstandard and poorly understood terminology (as attested to by the way 'warranted assertibility' has been misconstrued by subsequent commentators). My point is that Dewey's categories of assertion of judgments and affirmation of propositions constituted an attempt on his part to emphasize his view that even scientific inquiry is a goal directed activity exhibiting features in common with technological, economic, moral, prudential, and political deliberation. Justifiable change in state of belief is itself a species of rational decision making. The ends of inquiries focused on obtaining information to settle some question may differ from the goals of political, moral, economic, or prudential decision making. Likewise the options faced will differ. But cognitive decision making remains a species of decision making. And this observation suggests that a concern with theories of rational choice should be of interest to those who wish to articulate a pragmatist version of the belief-doubt model.

Presumably such an account of rational decision making would have to observe a neutrality with respect to the kind of decision problem under consideration. It must be an account of rational decision making which is applicable both to cognitive decision problems and to other kinds of decision making which differ from one another with respect to both their goals and the means and options deployed.

To be sure, no articulate account of rational decision making can avoid making some general assumptions about decision problems in general. So-called Bayesian decision theories, which promote maximization of expected utility, presuppose that the deliberating agent takes for granted that he or she has a certain set of options available to him or her and that if the agent were to become certain that he or she will implement one of these options, one of a given list of conjectures as to what the relevant 'outcome' of the implementation will be is true. According to the strict Bayesian view the agent is committed to a network of credal or subjective probability judgments concerning the truth of these conjectures conditional on the option being implemented and also to an evaluation of the hypotheses about outcomes representable by a utility function.[3]

potential answers to the question under investigation. This is the task of what Peirce called 'abduction.' Three minimal demands should be imposed on a potential answer: (a) Its truth should be consistent with what is taken for granted, (b) it should be a relevant answer to the question raised, and (c) whether it is to be accepted or rejected should be decidable through inquiry. Condition (a) indicates that potential answers or conjectures in Dewey's sense presuppose a distinction between certainties and conjectures in my sense.

[3] The utility function representation is to be understood in this setting as characterizing the agent's values and goals in the deliberation and not as representing the net of pleasure over pain of his or her desires. The utility function could represent moral,

I am not concerned for the present with the technical details of this Bayesian vision of rational decision making. Nor, for that matter, am I in favor of this Bayesian vision without substantial qualification.[4] My point in mentioning it is that it illustrates a feature of such weak conceptions of rational choice. There is not only a need for principles for evaluating the deliberator's options but also for conditions of coherence or consistency on the judgments as to which options are feasible, what the possible consequences of a given option are, the credal probability judgments and the utility judgments. The inquiring or deliberating agent is assumed to be drawing not only a distinction between what is certain and what is conjectural but finer grained discriminations among the conjectures with respect to credal probability and utility. We need to identify the constraints on rationality imposed on these judgments. To this extent, therefore, we need, in Alan Gibbard's words, an account of not only 'wise choices' but 'apt feelings.'[5]

We need to do more than this. The credal probability and utility judgments of agents, like the distinction between certainty and conjecture, are subject to change. A sophisticated version of the belief-doubt model of inquiry should provide an account of when and how judgments of credal probability and of value should be revised in the course of deliberation and inquiry. And this suggests that we should provide an account of a distinction between what is settled and what is doubtful in probability and utility judgment.

Thus, Charles Peirce was vociferously skeptical of the Bayesian idea that rational agents should have numerically definite credal probability judgments.[6] Unless such judgments are suitably grounded in knowledge of objective probabilities or chances, he thought they were doubtful. John Dewey insisted that doubt concerning what is for the best is a ubiquitous feature of our moral predicament and called for an extension of the belief-doubt model into questions of value.[7]

If one is going to give an account of rational decision making, rational full belief, rational probability judgment and rational value or utility judgment which offers any hope of allowing for a viable account of the budget of questions just thrown out for consideration, it is to be expected

political or cognitive valuations. The core principle involved is the principle of maximizing expected utility conditional on the act chosen. See Levi (1980, ch. 4) for a survey of core features of Bayesianism.

[4] For further elaboration of these qualifications, see Levi (1980 and 1986).

[5] Alan Gibbard (1990).

[6] See C. S. Peirce (1984–86, v. 2, pp. 98–102, v. 3, pp. 300–1) and Levi (1991a, pp. 99–103).

[7] See John Dewey and John H. Tufts (1932); and Levi (1986, ch. 1).

that the principles of rationality regulating these various propositional attitudes should be weak. The reason is that the account of change of view (whether it is change in what is counted as certain, what is judged probable or valuable) will depend on holding these principles of rationality fixed. Perhaps there is some deep sense in which no principles are to be held constant; but if one is seeking to give a systematic account of deliberation and inquiry, relative to that account some principles are going to be immune to revision. If the account is to be supple enough to provide a robust account of deliberation and inquiry, the fixed principles of coherent or consistent choice, belief, desire, etc. will have to be weak enough to accommodate a wide spectrum of potential changes in point of view. We may not be able to avoid some fixed principles, but they should be as weak as we can make them while still accommodating the demand for a systematic account.

To this extent, the view of rationality I am deploying departs from Gibbard's conception according to which 'to call something rational is to express one's acceptance of norms that permit it' (1990, p. 7). Granted that acceptance of norms of coherence or consistency controls the rational permissibility of 'choices' and 'feelings.' But in giving an account of the rational revision of norms through inquiry as Dewey sought to do, some sort of distinction needs to be made between the norms open to revision and those which are taken to characterize the rational conduct of inquiries concerned with such revision. Dewey thought, for example, that moral inquiry is just that kind of concern so that the moral principles which are the objects of criticism cannot be norms characterizing the rationality of the inquiry.[8]

The accounts of deliberation and inquiry offered by Peirce and Dewey were not intended to be fragments of sociology or psychology. Although Dewey insisted that certain biological and social conditions are necessary

[8] It may be pointed out that Dewey himself would have regarded the distinction I am making between norms of rationality and other norms as an untenable dualism. All norms are in some context or other objects of criticism and inquiry. But if our aim is to offer a systematic account of the rational conduct of inquiries concerned with the revision of beliefs, goals, values and other attitudes, the proposed accounts we offer will perforce draw a distinction between the norms which are held fixed in these accounts and those which are open to revision. I grant that disputes can arise as to which norms should qualify as the fixed norms of rationality. But any specific proposed account of inquiry acknowledges some distinction between norms which are fixed and norms which are open to revision. Rival proposals should, if they entertain the ambition of being systematic, share in common recognition of some sort of distinction between the fixed minimal norms of coherence or consistency and norms which are open to revision even if they differ concerning how the line is to be drawn. I propose to restrict the norms of rationality relative to an account of inquiry to the fixed minimal norms specified by the theory.

for inquiry to take place and, indeed, for well conducted inquiry to take place, good methods of inquiry are distinguished from bad and the characterization of norms of method is prescriptive.

I have suggested that the weak principles of rationality to be constructed could be construed as normative standards of rational health. Alternatively, they could be deployed by deliberating agents to evaluate their options, full beliefs, probability judgments and value judgments to ascertain whether they satisfy the requirements of a weak account of coherence and consistency. That is to say, the principles of rationality should be applicable in self criticism.

Peirce and Dewey thought of the prescriptions central to the belief-doubt model as available to the inquiring or deliberating agent for the purpose of self criticism in the context of deliberation or inquiry.[9] To be sure, such norms were also to serve as standards of rational health. But individuals well educated according to such standards were presumably to be trained to think for themselves - that is to say, to be in a position to bring the standards for rational health to bear in their own deliberations. In any case, I think an account of inquiry should be characterizable by norms available for nonvacuous self criticism. Hence, the standards of rationality relevant to the discussion of the belief-doubt model should be norms of this variety.

When used for self policing, the applicability of the principles should

[9] Neither Dewey nor Peirce explicitly claims this. Principles of reasoning are habits or leading principles or the like. The rhetoric, however, could often be read as blueprints for rational automata or as principles applicable in self criticism. Yet some passages suggest the latter reading fairly clearly. Thus Dewey writes: 'A postulate is also a stipulation. To engage in an inquiry is like entering into a contract. It commits the inquirer to certain conditions. A stipulation is a statement of conditions that are agreed to in the conduct of some affair. The stipulations are at first implicit in the undertaking of inquiry. As they are formally acknowledged (formulated), they become logical forms of various degrees of generality. They make definite what is involved in a demand. Every demand is a request, but not every request is a postulate. For a postulate involves the assumption of responsibilities. . . . On this account, postulates are not arbitrarily chosen. They present claims to be met in the sense in which a claim presents a title or has authority to receive due consideration' (Dewey, 1938, pp. 16-17). Dewey continues later to observe that when a specific person engages in inquiry, 'he is committed, in as far as his inquiry is genuinely such and not an insincere bluff, to stand by the results of similar inquiries by whomever conducted. 'Similar' in this phrase means inquiries that submit to the 'same conditions or postulates' (ibid., 18). The postulates Dewey is talking about here and which he regards as the terms of the contract an agent enters in undertaking an inquiry are clearly, for this reason, normative or prescriptive. As terms of a contract, they are intended to formulate prescriptions which the party to the contract endeavors to meet. When the undertakings are explicit, Dewey regards them as postulational. I take this to mean that postulates are principles the agent can explicitly recognize and use in evaluating the extent to which he is fulfilling his contract.

be nonvacuous in the sense that a nontrivial distinction may be made between feasible options which are admissible for choice and others which are not. If the principles of rational choice never eliminate any feasible option from the relevant set of feasible options, they fail to serve this function. It may still be possible to construe such principles as blueprints for designing rational automata or as principles for predicting behavior. But they will fail as standards of rational health for self critical agents and as principles for self policing.

The thesis I wish to advance is that this demand for nonvacuous self applicability entails an asymmetry between the first person perspective and the third person perspective which has no bearing on first person privileged access but which does pose a serious obstacle to viewing principles of rational choice designed to be nonvacuously applicable in self criticism as generalizations useful in prediction and explanation of human behavior. I shall argue that the asymmetry thesis does not hold if we rest content with viewing the normative principles of rationality as blueprints of conceptual, deliberative or rational health without expecting that agents use them in policing their own decisions. In that case, however, standards for rational health are blueprints for rational automata which simulate the behavior of rational agents but fail to employ the principles of rational choice to determine which feasible options are admissible. It is not clear to me that the classical pragmatists appreciated this point. As a consequence, the impression was given that emphasis on the explanatory uses of principles of rationality coheres well with the main tenets of pragmatism. In any event, many contemporary thinkers who identify themselves with pragmatism do not appear to be sensitive to the issue.

The asymmetry thesis is rich in consequences for contemporary conceptions of the mind and for theories of rational choice. In this essay, I can do no more than gesture towards what these consequences are. I shall rest content here in sharpening and defending the asymmetry.

A decision maker X engaged in deliberation needs to identify a roster of options X judges feasible for choice. Such judgments are crucial to any assessment by X of what it is rational for X to choose. X might recognize an option which, were it available to him, would be preferable to one he judges available to him. But if X judges the option unavailable to him, he is not required to judge it irrational of him to refuse to choose it rather than one of the options he recognizes as feasible. In applying criteria of rational choice to identify a set of options admissible for X to choose, X applies these criteria to a set of options which are feasible according to X.

If Sam, for example, is confronted with the choice of playing a piece by Chopin on the piano or surrendering the contents of his wallet and doubts that he has the ability to play the Chopin piece by his choice, he does not recognize playing the piece to be feasible for him. If Sam judged it feasible, he would prefer playing than paying. Failing to play would be irrational. But it is not irrational for him to pay if he cannot play.

Perhaps, the following objection will be raised: Unless Sam is certain that he is incapable of playing by choice, he has, from his point of view, the option of *trying* to play.

That may well be true. But trying to play is a different option from playing. To judge himself as having the option of playing, Sam must take for granted that his choice of playing is efficacious. He must be convinced that he will play if he chooses. If he were to have doubts about efficaciousness, he should not judge it feasible for him to play but at most to try to play. And whatever may be meant by 'trying to play' (e.g., making it objectively more probable that Sam will play), if trying to play is an option, Sam should be certain that his choosing this option will be efficacious - i.e., will render it objectively more probable that he will play. Being certain that the choice of an act is efficacious is a second necessary condition on judgments of feasibility.

Whether Sam judges that he can play the piano by choice or that he can only try, Sam is taking for granted that he has certain abilities (e.g., to play by choice). Sam's judgments of his objective abilities belong in Sam's state of full belief. Hence, Sam's state of full belief (and this holds true for any deliberating agent) must contain more than logical and other conceptual or a priori truths.

Abilities are duals of (sure fire) dispositions and, like sure fire dispositions, are relative to kinds of trials, experiments or initiating conditions. When a piece of sugar is alleged to have the sure fire disposition to dissolve in water, it is taken for granted that any water soluble thing dissolves if immersed in water. When a coin is alleged to have the ability to land heads on a toss (conditional on the coin's being tossed), it is taken for granted that anything possessing this ability lacks the sure fire disposition to fail to land heads on being tossed. Thus, if Sam has the ability to play Chopin by choice (i.e., conditional on deliberating) he lacks the sure fire disposition to fail to play Chopin by choice.

The relativity to kinds of trials or initiating conditions is of crucial importance. Sam may have the ability to play Chopin conditional on deliberating but at the very same time lack the ability to play Chopin by deliberating while suffering from an asthma attack. Sam could consistently be certain that he has the first ability and lacks the second. His conviction that he has the first ability may be necessary to his judging his choosing to play the piano to be a feasible option. But it is scarcely sufficient. If he also is certain that he is having an asthma attack while deliberating, it is not possible as far as Sam is concerned that he play the piano. That is to say, Sam is certain that he will not play the piano. Even though he has the ability at the time to play the piano conditional on deliberating and is deliberating, as far as he is concerned playing the piano is not optional for him. The extra information that he is suffering from an asthma attack precludes his coherently judging that playing is feasible for him. It is not irrational for him to surrender his wallet even though he would have wanted to play the piano rather than to pay were he facing that choice.[10]

Thus, Sam's full beliefs identify (1) what it is (objectively) possible for him to do through his deliberations (i.e., to choose to do) and (2) whether his choices are efficacious.

Given that Sam is certain that he has the objective ability to play Chopin through his choice and that his choices are efficacious, necessary conditions are satisfied for his playing Chopin to be feasible. Are these necessary conditions jointly sufficient? I think not. Feasibility presupposes that implementing the option is a 'serious possibility' - i.e., is consistent with the agent's state of full belief. If Sam is certain that he will not yield his wallet, paying is not possible as far as he is concerned. Once the matter is settled in this respect so that it is no longer consistent with what he takes for granted, feasibility is also precluded. That is to say, even if Sam would have preferred paying to playing (assuming paying was a serious possibility), it is not irrational for Sam to play given that paying is not a serious possibility relative to what he knows.

This third condition obtains regardless of how Sam came to be certain of this. Perhaps, it is because Sam has already decided to play. Perhaps, Sam's decision displayed weakness of will. He initially renounced paying while preferring to do so. His incontinence may be a form of irrationality.

[10] The assumptions which the deliberating agent makes concerning his abilities resemble in certain important respects the assumptions made when an inquirer judges that a stochastic experiment is to be implemented. If a die is about to be tossed once, the inquirer presupposes that exactly one of six kinds of outcome is about to occur. These six possible outcomes or points in the sample space represent abilities of the die to respond in these six ways on a toss. The die is also presupposed to have a sure fire disposition to land in exactly one of these ways on a toss. In deliberation, the agent makes analogous assumptions. He takes for granted that he has the ability to make true each of a variety of propositions through his deliberation (through his choice) and that he is constrained by deliberation to make exactly one of these propositions true. The space of objectively feasible options is like a sample space. See Levi (1986, ch. 4).

Yet it remains just as incoherent to continue to judge the option of playing as feasible. This point has, perhaps, more bite in cases where the agent is initially faced with more than two options. Jones may have interviewed three candidates for a job and may have decided (rationally or irrationally as the case may be) to reject the third candidate. As matters stand, the only available options are to hire the first or the second candidate. Given his decision to reject the third candidate and the efficaciousness of his choices, it is not epistemically or seriously possible that he choose the third candidate as far as he is concerned. Because hiring the third candidate is not a feasible option given Sam's convictions, rationality does not require that he take that option into account in determining what to do.

It may, perhaps, be objected that Sam can renege on his past decision. *If* reneging is an option for him and *if* he is not certain that he will not renege, the point is well taken. But given that Sam has chosen to reject the third candidate under the assumption of efficaciousness, he has ruled out reneging as a serious possibility. To be sure, Sam may subsequently change his mind and conclude that his initial decision is not efficacious after all. But as long as he fails to do so, he remains certain that he will not choose the rejected option. Consequently, in the context of his deliberation at the time, the rejected option is not a feasible option for him.

Thus, whether Sam coherently judges an option as feasible for him in the context of his current deliberation depends not only on his taking for granted assumptions about his abilities and efficaciousness but on his *not* taking for granted certain other claims but rather regarding them as serious possibilities. Sam's judgment of feasibility is dependent not only on what he knows but on his ignorance as well.

Suppose then that Sam judges that both playing the Chopin and giving up his wallet are optional for him. To determine what he should do rationally, Sam needs to identify his goals and values and how they together with his full beliefs and credal probability judgments determine which of the options he judges available to him are 'admissible' – i.e., not prohibited for choice. Thus, to apply his criteria of rationality to his problem, Sam needs to access information about his goals and values, his full beliefs and credal probability judgments and his criteria of rational choice and have enough computational capacity or logical omniscience to identify which of the options is admissible or, if both are, to reach this conclusion as well.

Sam may or may not manage these feats. Confusion, emotional disturbance, self deception and the like inhibit Sam's efforts to identify his values and beliefs. And if the structure of the problem is complex enough, he may lack the computational capacity to reach a definite conclusion on the basis of the information he has succeeded in eliciting. Thus, it is clear that criteria for rational choice can fail to be self applicable and often are. This is so whether 'Bayesian' or rival criteria are used provided they carry sufficient suppleness to reflect nuances of different types of decision problem. Such sophistication is always accompanied by the threat that decision problems will be confronted which call for more self awareness and computational capacity than the decision maker can muster. Hence, it is pointless, I think, to follow H. A. Simon and other devotees of 'bounded rationality' and seek to weaken the demands of rationality so as to guarantee that prescriptive rationality useful for self criticism coheres with the decision maker's abilities.

We may respond to these problems by acknowledging that principles of prescriptive rationality recommend conforming to their dictates insofar as the agent is capable of doing so while insisting that even when the agent is incapable, it is desirable to develop therapies, educational programs and technologies which enhance his capacity to conform better. We will rest satisfied with our prescriptive norms of rational choice as long as (i) they are applicable in a certain important category of sufficiently computationally undemanding cases free of emotional disturbance and (ii) therapies and technologies for enlarging the domain of applicability are available or are worth developing.

Suppose then that Sam is, indeed, in touch with his beliefs and desires, understands his principles of choice and has enough logical omniscience to identify the option (say, playing the piano) which is admissible according to his principles given his beliefs and desires.

If at the time Sam has figured all this out Sam takes for granted that he will choose rationally (i.e., choose an admissible option), Sam has the omniscience sufficient to conclude that he will choose to play the piano. That is to say, at that time, Sam is certain that he will play the piano and that he will not offer up his wallet.

But if that is Sam's view at the time he applies his principles of choice, from his point of view at that time, surrendering his wallet is not optional for him; for it is inconsistent with his state of full belief that he surrender his wallet - a point he can easily recognize.

There is no contradiction in this result. What it shows is that under the given assumptions Sam has only one option - the uniquely admissible option of playing the piano. Vacuity, not inconsistency, is the trouble. The set of feasible options and the set of admissible options coincide.

The argument illustrated by our example generalizes. If in addition to having the logical omniscience and self knowledge requisite to applying his principles of choice to identify a set of feasible options, the agent is convinced that he will restrict his choice to the admissible options, no inadmissible options are feasible. The principles of choice are applicable; but they are vacuously applicable. They cannot be used to reduce a feasible set to a proper subset of admissible options.

If we are driven to this conclusion, principles of rational choice become useless for the purposes of self policing of decisions.

It will not do to suggest that during the process of deliberation prior to figuring out which options are admissible, Sam has not ruled out the inadmissible options. As long as Sam has not identified the relevant values and beliefs and performed the requisite calculations, the principles of rational choice have not been applied. Only when all the information is in and the calculations have been made has a successful application of the principles of choice to determine admissibility been made. But at that point, the assumption that Sam will choose an admissible option precludes inadmissible options from being feasible. The set of admissible and the set of feasible options coincide. The application becomes vacuous.

Frederic Schick has, in effect, suggested in the face of this predicament that we jettison the idea that principles of rational choice are norms for self criticism.[11] He favors the view that they are principles for the prediction and explanation of choices. In so doing, he acknowledges the main point I mean to press - to wit, that seeing such principles as explanatory and predictive coheres poorly with using them nonvacuously as norms for self criticism.

The alternative view, which I favor, seeks to preserve the status of principles of rationality as nonvacuous norms for self criticism. But if the deliberating agent takes for granted the 'smugness assumption' which asserts that the deliberating agent will choose rationally, the principles of rationality cannot be nonvacuously self applicable. At best, they can be vacuously applicable.

On the other hand, if the smugness assumption is abandoned, the principles of rationality remain self applicable; but now they can be nonvacuous. From the perspective I favor, therefore, the smugness assumption ought to be abandoned.

The argument does not prevent the agent X from predicting that agent Y will choose an admissible option or that X himself will choose ratio-

[11] See F. Schick (1979, p. 243). See also Levi (1986, ch. 4, sec. 3) for a further discussion of views on this issue.

nally in some future deliberation. Prediction is precluded only for X predicting his own rational choice in the current context of deliberation. The asymmetry between the first person and third person perspectives implied thereby does not imply a privileged access. It derives from the assumption that canons of rationality are to be used in self criticism in the context of deliberation - that is, in identifying what is to be chosen. I shall return to this point shortly.

Perhaps, it will be argued that even if the decision maker cannot take for granted or be certain that he will choose rationally, he can at least judge that it is probable that he will choose rationally.

But our assumptions about necessary conditions for judgments of feasibility preclude this as well. Indeed, it can be argued that the decision maker cannot make any coherent judgments of credal probability relevant to action concerning what he will choose in the current deliberation.

Consider Sam again. Suppose Sam prefers playing Chopin to yielding his wallet. He is offered a bet as to whether he will play or yield his wallet where he wins $100 if he plays and nothing if he yields his wallet. If he has credal probabilities for what he will choose, the amount he will be willing to pay will be controlled by his credal probabilities. It is clear that Sam should be prepared to pay $100 for the bet. To see this, observe that Sam will clearly prefer playing and taking the bet for any fee less than $100 to playing and refusing to pay that fee and will prefer yielding the wallet and refusing the bet for any fee to yielding the wallet and accepting the bet for a fee. So we need to compare playing and taking the bet with paying and refusing. Given that Sam prefers playing to paying, he should prefer playing and taking the bet to paying and refusing. Moreover he should do so for any fee short of $100. Thus, the 'fair betting rate' for the hypothesis that he will play is 1 and, so it seems, Sam is certain that he will play. Hence, that option and that option alone is feasible for Sam. The admissible and the feasible set coincide.

The upshot is that if Sam is to deploy criteria for choice to determine what he should do, he must not make any judgments as to the probability as to what he will do. A fortiori, he should not make any judgments as to the probability that he will choose rationally. To be an agent crowds out being a predictor.

Suppose that Sam is convinced that he is able to play Chopin through his own choice but is not able to play Chopin through his choice while suffering an asthma attack. I have said that Sam can consistently be convinced of the truth of both claims. Suppose also that Sam is in doubt as to whether he is suffering an asthma attack. Then he must be in doubt as to whether his choice is efficacious and, hence, cannot judge choosing to play to be feasible. (Trying to play may be feasible; but that is a different matter.) We shall assume that Sam is certain that no asthma attack will occur.

These remarks apply, as I have said, to Sam's judgments (or, for that matter, to the judgments of any deliberating agent X) at that stage in deliberation where the agent has identified his values, convictions and options sufficiently to apply the principles of rationality to the evaluation of the admissibility of these options. They do not apply to Sam's evaluation of the rationality of Y's choices or to the rationality of Sam's choices in some other future context of deliberation. Sam can attempt to identify Y's beliefs and values and Y's judgments of feasibility and then apply the principles of rationality to determine how someone in Y's position should choose and still make a prediction as to what Y will choose. If Sam regards Y to be coherent in his judgments, he will not attribute to Y a prediction as to how Y will choose but he can coherently make a prediction of his own.

Since Y can be Sam himself in some future deliberation, we can consider Sam's situation prior to facing the decision whether to play or pay. Suppose Sam is convinced before the Moment of Truth that he will face a decision. There is nothing in our argument which precludes Sam predicting that he will choose an admissible option - i.e., choose rationally. But since he is convinced that he will face a decision where he will choose, Sam is also committed prior to the Moment of Truth to the view that at the Moment of Truth he will cease taking for granted that he will choose rationally.

Moreover, prior to the Moment of Truth, Sam may be in a position to predict what his values and beliefs will be at the Moment of Truth and to determine which of the options he will judge feasible at the Moment of Truth are admissible. Given his conviction prior to the Moment of Truth that he will choose rationally, he will be committed to a prediction as to which act Sam will choose at the Moment of Truth. If there is exactly one admissible option (playing Chopin), the prediction should be that that option will be chosen. If there are several, the prediction will be that one of those admissible options will be chosen.

With the passage of time, Sam arrives at the Moment of Truth. If Sam's prior prediction that Sam will face a decision at that point is borne out, Sam will cease being convinced that he will choose rationally. Moreover, he will cease being convinced as to which of the specific feasible options he will choose.

Thus, Sam will have modified his state of full belief by 'contraction' -

i.e., by giving up some full beliefs. According to the belief-doubt model of inquiry, changes in states of full belief call for justification. What can justify the change in this case?

Observe, at the outset, that if Sam does not give up the prediction that he will choose rationally and, more specifically, that he will play Chopin, then, if he is to preserve coherence, he will have to give up the prediction that he faces a choice between playing and paying at the Moment of Truth. So he will have to give up some prior assumption at the Moment of Truth if coherence is to be preserved. That is coerced by the demands of coherence. The problem of justification concerns what to give up.

A strong case can be made for giving up the prediction that he will choose rationally. Suppose that Sam instead abandons the prediction that he will face a choice between playing and paying. I am supposing that no new information has been obtained by Sam during the interim between his initial predictions and the Moment of Truth which would offer independent reasons for giving up this prediction, so that the only reason for abandoning the prediction that he will face a choice is to preserve coherence and preserve the prediction that he will choose rationally between playing and paying. This strategy is self defeating. Sam can choose rationally between playing and paying only if he faces a choice between these two options. If the only option he faces is playing, he cannot choose rationally between these two options. He may, perhaps, be said to choose in a degenerate sense and, indeed, again in a degenerate sense, to choose rationally. But he is not choosing rationally in the respect in which it was predicted that he would. The upshot is that Sam must abandon the prediction that he will choose rationally between playing and paying whether he retains or abandons the prediction that he will choose between playing and paying. Under the circumstances, abandoning the prediction that he will choose between playing and paying entails a gratuitous reduction in the information available to Sam in his state of full belief. In the absence of an independent justification, he should not do it.

It may, however, be objected that Sam will also have to abandon the prediction that he will play Chopin. Even though Sam cannot claim at the Moment of Truth that he will choose playing over paying, he can at least predict that he will play. The objection is that in retaining the claim that he will face a choice between playing and paying, he abandons this prediction as well. Thus, it may appear that giving up the prediction that Sam faces a decision at the Moment of Truth may not be gratuitous after all.

This objection, however, is not compelling. We are supposing that Sam would not have been convinced that he would play Chopin at the Moment of Truth had he not predicted that he would choose rationally. Since the prediction that he will choose rationally is going to have to be given up anyhow, the prediction that he will play incurs no further loss of explanatory or informational value. Sam is justified in giving up the prediction that he will play.[12]

This would not be true if prior to the Moment of Truth Sam was certain that he would play for other reasons. For example, Sam might have been convinced before the Moment of Truth that he would be incapable of surrendering his wallet because he had already been robbed. But given that information, he should not predict beforehand that he would choose rationally between playing and paying or, indeed, that he would choose between these options at all. And when the Moment of Truth comes, he should not abandon his prediction that he will play unless he has independent good reason for doing so.

When, however, the sole basis for the prediction that he will play is that he will choose between playing and paying rationally (together with the assumptions leading to the conclusion that Sam prefers playing to paying), giving up the rationality assumption warrants giving up the prediction that he will play. And there is no need to justify giving up the prediction that he will choose between playing and paying rationally. Given the other background assumptions, coherence in judgments of feasibility requires doing so at the Moment of Truth.

The cogency of these arguments depends critically on my contention that norms of rational choice should be nonvacuously applicable by the decision maker in policing his deliberations. If one insists on regarding such norms as functioning primarily as principles of an explanatory and predictive theory, the argument fails. The idea that belief-desire models of human behavior controlled, at least in idealization, by principles of rationality constitute an explanatory and predictive theory of human behavior has been widely supported. The arguments I have adduced do not refute this assumption unless the claim that principles of rationality are to be used as norms for self policing is endorsed.

I question the status of principles of rationality as explanatory laws on quite independent grounds. It is well known that we lack the computational capacity, memory and psychic stability to satisfy principles of rationality except in limiting cases. To finesse this difficulty, it is often

[12] Appeal is being made here to accounts of contraction which recommend violations of the so-called Recovery Postulate under the conditions envisaged in the text. The Recovery Postulate is discussed and defended in Peter Gärdenfors (1988, ch. 3.4). Violation of this postulate is defended in Levi (1991b, ch. 4, sec. 5).

said that models of rational behavior are 'idealizations' just as theories of ideal gases are, so that they have the kind of explanatory force which is found in the natural sciences. I do not think the analogy is apt. Some real gases approximate the conditions of ideal gases. Human agents never remotely come close to satisfying conditions for ideal rationality if for no other reason than that we lack logical omniscience. To be sure, humans satisfy the conditions well enough in some cases to make decisions meeting the requirements of rationality; but even in those cases, they fail miserably in satisfying all the conditions of rationality. Moreover, modifying the ideal theory so as to improve upon the approximation is always regarded as worthwhile in science. If principles of rationality are intended to be explanatory, we should seek to replace rational explanation by a better theory. Consequently we should abandon the initial principles of rationality as an explanatory theory except perhaps in the way in which discarded scientific theories are used instrumentalistically as crude first approximations. I contend that we should not jettison theories of rationality so easily as that. We know very well that our full beliefs as to what is true are full of logical inconsistencies but we continue or should continue to attach importance to the desirability of removing inconsistencies when we recognize them. Rather than abandoning models of rationality, we should seek instead to devise techniques and therapies which enhance our capacities to do better. In this respect, models of rationality bear a closer resemblance to models of health and mental health. They are, for this reason, normative rather than explanatory and predictive.[13]

Even if this is conceded, however, the conclusions I have been advancing may be resisted. Perhaps models of rationality are designs for better human agents. But we may make such designs without deploying them in self policing. We can design automata to behave in desirable ways without insuring that the automata use the principles of design to police their own behavior. Normativity alone will not bring the conclusions I

[13] Akeel Bilgrami (1991) mounts a convincing attack on theories of so-called wide content, direct reference and the like because of their inability to show how content so construed can be deployed in the explanation of human behavior. The weak link in Bilgrami's argument, so it seems to me, is his assumption (shared with many of those he criticizes) that the primary function of appeals to beliefs and desires is in explanation of behavior. However, it seems to me that this reservation with Bilgrami's argument does little to damage it. Even if principles of rationality are norms rather than laws explaining behavior, we should want to claim that were we rational agents completely satisfying the dictates of such principles, self criticism would be otiose, rational angels would be rational automata and our behavior would be explainable by these principles. Indeed, were this not the case, we would regard the principles of rationality as somehow defective as norms for use in self criticism.

have been defending. Appeal must be made to the use of the norms in self policing of deliberation. Thus, those who resist the conclusions I am advancing must insist on external policing to the exclusion of internal policing.

The point is well taken. External policing can take the form of deploying norms of rationality as blueprints for rational automata. There is no clash between using principles of rationality for explanatory and predictive purposes, on the one hand, and using them prescriptively for designing rationally acceptable conduct. Tension arises only if, in addition to using them for external policing, one seeks to use them for internal policing. In that event, the blueprints can no longer be for rational *automata.* Agents will satisfy the requirements for rational health only if they apply the principles of choice to evaluate their options. But, in that case, neither they nor we, the outside agents, can regard them as predicting their own choices. Rational automata can predict their own choices. Rational agents cannot.

The tedious mental gymnastics of the paragraphs dealing with Sam before and at the Moment of Truth are intended to undermine the impression one might receive that the asymmetry between the first personal point of view and the third personal point of view smacks of mystery or irrationality. Once the roots of the asymmetry in the demand that principles of rational choice be nonvacuously self applicable are recognized, there should be neither mystery nor irrationality.

Mysterious or not, the implications of the approach I am pressing are far reaching for accounts of rational decision making, rational probability judgment and rational value judgment. I have discussed some of these ramifications elsewhere. Space does not permit elaboration on them here.[14]

I declared my interest in the accounts of rational choice as deriving from my interest in a systematic development of the pragmatist belief-doubt model of inquiry as a prescriptive account. On the basis of this interest, I defended the idea that principles of rationality should be relatively weak, context independent principles, applicable in a wide variety of contexts and that they should be relevant principles for assessing both wise choices and apt feelings.

Is there any element in the pragmatist approach which might argue for or against insisting on construing the prescriptions of rationality as nonvacuously applicable in self criticism and deliberation rather than as recipes for designing rational automata? I have not been able to find any

[14] See Levi (1987, pp. 193-211, 1991c, ch. 4, and 1992, pp. 1-20).

clear indication in the writings of the classical pragmatists that they recognized the need to consider the issue. Authors who in one way or another have intimated their pragmatist sympathies subsequently have felt free to adopt positions which, if the argument of this essay is right, abandon the use of principles of rationality for purposes of self criticism.

Consequently, my appeal to an interest in articulating a pragmatist account of problem solving inquiry as a means for identifying a conception of rationality has exhausted its resources without fully settling the question. If one appeals to Dewey's famous interest in education aimed at training students in the methods of well conducted problem solving inquiry, one will still face the issue of deciding whether the training will require inculcating a capacity for self criticism oı whether it is enough to produce well trained seals. I have no doubt that Dewey would have opted for the former over the latter alternative. But appeals to authority will not settle the issue. Still if there is any remnant of the Kantian view of autonomy worth preserving from a pragmatist perspective, it is to be found in the nonvacuous applicability of standards of rationality in self criticism.

Opponents and proponents of preserving this remnant should recognize, however, the presence of the tension between the explanatory use of principles of rationality and their use as norms for self criticism. The tension is severe and fraught with consequences deserving serious philosophical reflection.

REFERENCES

Bilgrami, A. (1991), *Belief and Meaning*, Oxford: Blackwell.

Dewey, J. (1938), *Logic: The Theory of Inquiry*, New York: Holt.

(1941), "Propositions, Assertibility and Truth," reprinted in *Dewey and His Critics*, ed. by S. Morgenbesser, New York: Hackett, 1977, 265-82.

Dewey, J. and Tufts, J. H. (1932), *Ethics*, New York: Holt.

Gärdenfors, P. (1988), *Knowledge in Flux*, Cambridge, MA.: MIT Press.

Gibbard, A. (1990), *Wise Choices, Apt Feelings*, Cambridge, MA.: Harvard University Press.

Levi, I. (1980), *The Enterprise of Knowledge*, Cambridge, MA.: MIT Press, ch. 4.

(1983), "Doubt, Context and Inquiry," in *How Many Questions*, ed. by L. Cauman et al., New York: Hackett, 25-34.

(1986), *Hard Choices*, Cambridge: Cambridge University Press.

(1987), "The Demons of Decision," *The Monist 70*, 193-211.

(1991a), "Chance," in *Philosophical Topics: Philosophy of Science*, ed. by L. Nissen, vol. 18, New York: Hackett, 95-121.

(1991b), *The Fixation of Belief and Its Undoing*, Cambridge: Cambridge University Press.

(1991c), "Consequentialism and Sequential Choice," in *Foundations of Decision Theory*, ed. by M. Bacharach and S. Hurley, Oxford: Blackwell, ch. 4.

(1992), "Feasibility," in *Knowledge, Belief and Strategic Interaction*, ed. by C. Bicchieri and M. L. Dalla Chiara, Cambridge: Cambridge University Press, 1-20.

Peirce, C. S. (1984-86), in *The Writings of Charles S. Peirce*, ed. by M. Fisch et al., Indianapolis: Indiana University Press, vol. 2, 98-102, and vol. 3, 300-301.

Schick, F. (1979), "Self Knowledge, Uncertainty and Choice," *British Journal for the Philosophy of Science 30*, 235-52.