Rational Fools: A Critique of the Behavioral Foundations of Economic Theory

Amartya K. Sen

Philosophy and Public Affairs, Vol. 6, No. 4 (Summer, 1977), 317-344.

Stable URL:
http://links.jstor.org/sici?sici=0048-3915%28197722%296%3A4%3C317%3ARFACOT%3E2.0.CO%3B2-Z

Philosophy and Public Affairs is currently published by Princeton University Press.

AMARTYA K. SEN

Rational Fools: A Critique of the Behavioral Foundations of Economic Theory

I

In his *Mathematical Psychics*, published in 1881, Edgeworth asserted that "the first principle of Economics is that every agent is actuated only by self-interest."[1] This view of man has been a persistent one in economic models, and the nature of economic theory seems to have been much influenced by this basic premise. In this essay I would like to examine some of the problems that have arisen from this conception of human beings.

I should mention that Edgeworth himself was quite aware that this so-called first principle of Economics was not a particularly realistic one. Indeed, he felt that "the concrete nineteenth century man is for the most part an impure egoist, a mixed utilitarian."[2] This raises the interesting question as to why Edgeworth spent so much of his time and talent in developing a line of inquiry the first principle of which he believed to be false. The issue is not why abstractions should be em-

This Herbert Spencer Lecture, delivered at Oxford University in October 1976, will appear in *Scientific Models and Man*, ed. H. Harris (forthcoming 1978) and is printed here by kind permission of Oxford University Press. For helpful comments on an earlier version, I am grateful to the Editors of this journal, and to Åke Andersson, Isaiah Berlin, Frank Hahn, Martin Hollis, Janos Kornai, Derek Parfit, Christopher Peacocke, and Tibor Scitovsky.

1. F.Y. Edgeworth, *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences* (London, 1881), p. 16.

2. Edgeworth (1881), p. 104. In fact, he went on to make some interesting remarks on the results of "impure" egoism, admitting an element of sympathy for each other. The remarks have been investigated and analyzed by David Collard, "Edgeworth's Propositions on Altruism," *Economic Journal* 85 (1975).

ployed in pursuing general economic questions—the nature of the inquiry makes this inevitable—but why would one choose an assumption which he himself believed to be not merely inaccurate in detail but fundamentally mistaken? As we shall see, this question is of continuing interest to modern economics as well.

Part of the answer, as far as Edgeworth was concerned, undoubtedly lay in the fact that he did not think the assumption to be fundamentally mistaken in the *particular* types of activities to which he applied what he called "economical calculus": (i) war and (ii) contract. "Admitting that there exists in the higher parts of human nature a tendency towards and feeling after utilitarian institutions," he asked the rhetorical question: "could we seriously suppose that these moral considerations were relevant to war and trade; could eradicate the controlless core of human selfishness, or exercise an appreciable force in comparison with the impulse of self-interest."[3] He interpreted Sidgwick to have dispelled the "illusion" that "the interest of all is the interest of each," noting that Sidgwick found the "two supreme principles—Egoism and Utilitarianism" to be "irreconcilable, unless indeed by religion." "It is far from the spirit of the philosophy of pleasure to deprecate the importance of religion," wrote Edgeworth, "but in the present inquiry, and dealing with the lower elements of human nature, we should have to seek a more obvious transition, a more earthy passage, from the principle of self-interest to the principle, or at least the practice, of utilitarianism."[4]

Notice that the context of the debate is important to this argument. Edgeworth felt that he had established the acceptability of "egoism" as the fundamental behavioral assumption for his particular inquiry by demolishing the acceptability of "utilitarianism" as a description of actual behavior. Utilitarianism is, of course, far from being the only non-egoistic approach. Furthermore, between the claims of oneself and the claims of all lie the claims of a variety of groups—for example, families, friends, local communities, peer groups, and economic and social classes. The concepts of family responsibility, business ethics, class consciousness, and so on, relate to these intermediate areas of concern, and the dismissal of utilitarianism as a descriptive theory

3. Edgeworth, p. 52.
4. Ibid., pp. 52–53.

of behavior does not leave us with egoism as the only alternative. The relevance of some of these considerations to the economics of negotiations and contracts would be difficult to deny.

It must be noted that Edgeworth's query about the outcome of economic contact between purely self-seeking individuals had the merit of being immediately relevant to an abstract enquiry that had gone on for more than a hundred years already, and which was much discussed in debates involving Herbert Spencer, Henry Sidgwick, and other leading thinkers of the period. Two years before Edgeworth's *Mathematical Psychics* appeared, Herbert Spencer had published his elaborate analysis of the relation between egoism and altruism in *The Data of Ethics*. He had arrived at the comforting–if somewhat unclear –conclusion that "general happiness is to be achieved mainly through the adequate pursuit of their own happinesses by individuals; while, reciprocally, the happiness of individuals are to be achieved in part by their pursuit of the general happiness."[5] In the context of this relatively abstract enquiry, Edgeworth's tight economic analysis, based on a well-defined model of contracts between two self-seeking individuals, or between two types of (identical) self-seeking individuals, gave a clear answer to an old hypothetical question.

It appeared that in Edgeworth's model, based on egoistic behavior, there was a remarkable correspondence between exchange equilibria in competitive markets and what in modern economic terms is called "the core" of the economy. An outcome is said to be in "the core" of the economy if and only if it fulfills a set of conditions of unimprovability. These conditions, roughly speaking, are that not only is it the case that no one could be made better off without making somebody else worse off (the situation is what is called a "Pareto optimum"), but also that no one is worse off than he would be without trade, and that no coalition of individuals, by altering the trade among themselves, could on their own improve their own lot. Edgeworth showed that given certain general assumptions, any equilibrium that can emerge in a competitive market must satisfy these conditions and be in "the core." Thus, in Edgeworth's model the competitive market equilibria are, in this sense, undominated by any feasible alternative

5. H. Spencer, *The Data of Ethics* (London, 1879; extended edition, 1887), p. 238.

arrangement, given the initial distribution of endowments. More surprising in some ways was the converse result that if the number of individuals of each type were increased without limit, the core (representing such undominated outcomes) would shrink towards the set of competitive equilibria; that is, the core would not be much more extensive than the set of competitive equilibria. This pair of results has been much elaborated and extended in the recent literature on general equilibrium with similar models and with essentially the same behavioral assumptions.[6]

Being in the core, however, is not as such a momentous achievement from the point of view of social welfare. A person who starts off ill-endowed may stay poor and deprived even after the transactions, and if being in the core is all that competition offers, the propertyless person may be forgiven for not regarding this achievement as a "big deal." Edgeworth took some note of this by considering the problem of choice between different competitive equilibria. He observed that for the utilitarian good society, "competition requires to be supplemented by arbitration, and the basis of arbitration between self-interested contractors is the greatest possible sum-total utility."[7] Into the institutional aspects of such arbitration and the far-reaching implications of it for the distribution of property ownership, Edgeworth did not really enter, despite superficial appearance to the contrary. On the basis of the achievement of competition, however limited, Edgeworth felt entitled to be "biassed to a more conservative caution in reform." In calculating "the utility of pre-utilitarian institutions," Edgeworth felt impressed "with a view of Nature, not, as in the picture left by Mill, all bad, but a first approximation to the best."[8]

I am not concerned in this essay with examining whether the approximation is a rather remote one. (This I do believe to be the case even within the structure of assumptions used by Edgeworth, but it is not central to the subject of this paper.) I am concerned here with the view of man which forms part of Edgeworth's analysis and

6. See, especially, K.J. Arrow and F.H. Hahn, *General Competitive Analysis* (San Francisco, 1971).
7. Edgeworth, p. 56.
8. Ibid., p. 82.

survives more or less intact in much of modern economic theory. The view is, of course, a stylized one and geared specifically to tackling a relatively abstract dispute with which Spencer, Sidgwick, and several other leading contemporary thinkers were much concerned–namely, in what sense and to what extent would egoistic behavior achieve general good? Whether or not egoistic behavior is an accurate assumption in reality does not, of course, have any bearing on the accuracy of Edgeworth's answer to the question posed. Within the structure of a limited economic model it provided a clear-cut response to the abstract query about egoism and general good.

This particular debate has gone on for a long time and continues to provide motivation for many recent exercises in economic theory today. The limited nature of the query has had a decisive influence on the choice of economic models and the conception of human beings in them. In their distinguished text on general equilibrium theory, Arrow and Hahn state (pp. vi–vii):

> There is by now a long and fairly imposing line of economists from Adam Smith to the present who have sought to show that a decentralized economy motivated by self-interest and guided by price signals would be compatible with a coherent disposition of economic resources that could be regarded, in a well-defined sense, as superior to a large class of possible alternative dispositions. Moreover, the price signals would operate in a way to establish this degree of coherence. It is important to understand how surprising this claim must be to anyone not exposed to the tradition. The immediate "common sense" answer to the question "What will an economy motivated by individual greed and controlled by a very large number of different agents look like?" is probably: There will be chaos. That quite a different answer has long been claimed true and has indeed permeated the economic thinking of a large number of people who are in no way economists is itself sufficient ground for investigating it seriously. The proposition having been put forward and very seriously entertained, it is important to know not only whether it *is* true, but whether it *could* be true. A good deal of what follows is concerned with this last question, which seems to us to have considerable claims on the attention of economists.

The primary concern here is not with the relation of postulated models to the real economic world, but with the accuracy of answers to well-defined questions posed with preselected assumptions which severely constrain the nature of the models that can be admitted into the analysis. A specific concept of man is ingrained in the question itself, and there is no freedom to depart from this conception so long as one is engaged in answering this question. The nature of man in these current economic models continues, then, to reflect the particular formulation of certain general philosophical questions posed in the past. The realism of the chosen conception of man is simply not a part of this inquiry.

II

There is another nonempirical—and possibly simpler—reason why the conception of man in economic models tends to be that of a self-seeking egoist. It is possible to define a person's interests in such a way that no matter what he does he can be seen to be furthering his own interests in every isolated act of choice.[9] While formalized relatively recently in the context of the theory of revealed preference, this approach is of respectable antiquity, and Joseph Butler was already arguing against it in the Rolls Chapel two and a half centuries ago.[10] The reduction of man to a self-seeking animal depends in this approach on careful definition. If you are observed to choose x rejecting y, you are declared to have "revealed" a preference for x over y. Your personal utility is then defined as simply a numerical representation of this "preference," assigning a higher utility to a "preferred" alternative. With this set of definitions you can hardly escape maximizing your own utility, except through inconsistency. Of course, if you choose x and reject y on one occasion and then promptly proceed to do the exact opposite, you can prevent the revealed preference theorist

9. If a person's actions today affect his well-being in the future, then under this approach his future interests must be defined in terms of the way they are *assessed today*. In general, there is no reason to presume that the future interests as assessed today will coincide with those interests as assessed in the future. This adds an additional dimension to the problem, and I am grateful to Derek Parfit for convincing me of the conceptual importance of this question.

10. J. Butler, *Fifteen Sermons Preached at the Rolls Chapel* (London, 1726); see also T. Nagel, *The Possibility of Altruism* (Oxford, 1970), p. 81.

from assigning a preference ordering to you, thereby restraining him from stamping a utility function on you which you must be seen to be maximizing. He will then have to conclude that either you are inconsistent or your preferences are changing. You can frustrate the revealed-preference theorist through more sophisticated inconsistencies as well.[11] But if you are consistent, then no matter whether you are a single-minded egoist or a raving altruist or a class conscious militant, you will appear to be maximizing your own utility in this enchanted world of definitions. Borrowing from the terminology used in connection with taxation, if the Arrow-Hahn justification of the assumption of egoism amounts to an *avoidance* of the issue, the revealed preference approach looks more like a robust piece of *evasion.*

This approach of definitional egoism sometimes goes under the name of rational choice, and it involves nothing other than internal consistency. A person's choices are considered "rational" in this approach if and only if these choices can *all* be explained in terms of some preference relation consistent with the revealed preference definition, that is, if all his choices can be explained as the choosing of "most preferred" alternatives with respect to a postulated preference relation.[12] The rationale of this approach seems to be based on the idea that the only way of understanding a person's real preference is to examine his actual choices, and there is no choice-independent way of understanding someone's attitude towards alternatives. (This view, by the way, is not confined to economists only. When, many years ago, I had to take my qualifying examination in English Literature at Calcutta University, one of the questions we had to answer concerning *A Midsummer Night's Dream* was: Compare the characters of Hermia and Helena. Whom would you choose?)

I have tried to demonstrate elsewhere that once we eschew the curious definitions of preference and welfare, this approach presumes both too little and too much: too little because there are non-choice sources of information on preference and welfare as these terms are

11. See H.S. Houthakker, "Revealed Preference and the Utility Function," *Economica* 17 (1950); P.A. Samuelson, "The Problem of Integrability in Utility Theory," *Economica* 17 (1950).

12. For the main analytical results, see M.K. Richter, "Rational Choice," *Preference, Utility and Demand Theory,* ed. J.S. Chipman et al. (New York, 1971).

usually understood, and too much because choice may reflect a compromise among a variety of considerations of which personal welfare may be just one.[13]

The complex psychological issues underlying choice have recently been forcefully brought out by a number of penetrating studies dealing with consumer decisions[14] and production activities.[15] It is very much an open question as to whether these behavioral characteristics can be at all captured within the formal limits of consistent choice on which the welfare-maximization approach depends.[16]

III

Paul Samuelson has noted that many economists would "separate economics from sociology upon the basis of rational or irrational behavior, where these terms are defined in the penumbra of utility

13. A.K. Sen, "Behaviour and the Concept of Preference," *Economica* 40 (1973). See also S. Körner's important recent study, *Experience and Conduct* (Cambridge, 1971). Also T. Schwartz, "Von Wright's Theory of Human Welfare: A Critique," forthcoming in P.A. Schlipp, ed., *The Philosophy of Georg Henrik von Wright*; T. Majumdar, "The Concept of Man in Political Economy and Economics," mimeographed (Jawaharlal Nehru University, New Delhi, 1976); and F. Schick, "Rationality and Sociality," mimeographed (Rutgers University, Philosophy of Science Association, 1976).

14. See T. Scitovsky, *The Joyless Economy: An Inquiry into Human Satisfaction and Consumer Dissatisfaction* (London and New York, 1976). See also the general critique of the assumption of "rational" consumer behavior by J. Kornai, *Anti-Equilibrium* (Amsterdam and London, 1971), chap. 11; and the literature on "psychological choice models," in particular, D. McFadden, "Economic Applications of Psychological Choice Models" (presented at the Third World Econometric Congress, August 1975).

15. See H. Liebenstein, "Allocative Efficiency vs. x-Efficiency," *American Economic Review* 56 (1966). Also critiques of the traditional assumption of profit maximization in *business* behavior, particularly W.J. Baumol, *Business Behavior, Value and Growth* (New York, 1959); R. Marris, *The Economic Theory of Managerial Capitalism* (London, 1964); O. Williamson, *The Economics of Discretionary Behavior* (Chicago, 1967); and A. Silberston, "Price Behaviour of Firms," *Economic Journal* 80 (1970), reprinted in Royal Economic Society, *Surveys of Applied Economics*, vol. 1 (London, 1973).

16. On the required conditions of consistency for viewing choice in terms of a binary relation, see my "Choice Functions and Revealed Preference," *Review of Economic Studies* 38 (1971); H.G. Herzberger, "Ordinal Preference and Rational Choice," *Econometrica* 41 (1973); K. Suzumura, "Rational Choice and Revealed Preference," *Review of Economic Studies* 43 (1976); S. Kanger, "Choice Based on Preference," mimeographed (Uppsala University, 1976).

theory."[17] This view might well be resented, for good reasons, by sociologists, but the cross that economists have to bear in this view of the dichotomy can be seen if we note that the approach of "rational behavior," as it is typically interpreted, leads to a remarkably mute theory. Behavior, it appears, is to be "explained in terms of preferences, which are in turn defined only by behavior." Not surprisingly, excursions into circularities have been frequent. Nevertheless, Samuelson is undoubtedly right in asserting that the theory "is not in a technical sense *meaningless*."[18] The reason is quite simple. As we have already discussed, the approach does impose the requirement of internal consistency of observed choice, and this might well be refuted by actual observations, making the theory "meaningful" in the sense in which Samuelson's statement is intended.

The requirement of consistency does have surprising cutting power. Various general characteristics of demand relations can be derived from it. But in the present context, the main issue is the possibility of using the consistency requirement for actual *testing*. Samuelson specifies the need for "ideal observational conditions" for the implications of the approach to be "refuted or verified." This is not, however, easy to satisfy since, on the one hand, our love of variety makes it illegitimate to consider individual acts of choice as the proper units (rather than *sequences* of choices) while, on the other hand, lapse of time makes it difficult to distinguish between inconsistencies and changing tastes. There have, in fact, been very few systematic attempts at testing the consistency of people's day-to-day behavior, even though there have been interesting and useful contrived experiments on people's reactions to uncertainty under laboratory conditions. What counts as admissible evidence remains unsettled. If today you were to poll economists of different schools, you would almost certainly find the coexistence of beliefs (i) that the rational behavior theory is unfalsifiable, (ii) that it is falsifiable and so far unfalsified, and (iii) that it is falsifiable and indeed patently false.[19]

17. P.A. Samuelson, *The Foundation of Economics* (Cambridge, Mass., 1955), p. 90.

18. Ibid., p. 91.

19. The recent philosophical critiques of rational behavior theory include, among others, M. Hollis and E.J. Nell, *Rational Economic Man* (Cambridge, 1975); S. Wong, "On the Consistency and Completeness of Paul Samuelson's

However, for my purposes here this is not the central issue. Even if the required consistency were seen to obtain, it would still leave the question of egoism unresolved except in the purely definitional sense, as I have already noted. A consistent chooser can have any degree of egoism that we care to specify. It is, of course, true that in the special case of pure consumer choice over private goods, the revealed preference theorist tries to relate the person's "preference" or "utility" to his *own* bundle of commodities. This restriction arises, however, not from any guarantee that he is concerned only with his own interests, but from the fact that his own consumption bundle—or that of his family—is the only bundle over which he has direct *control* in his acts of choice. The question of egoism remains completely open.

I believe the question also requires a clearer formulation than it tends to receive, and to this question I shall now turn.

IV

As we consider departures from "unsympathetic isolation abstractly assumed in Economics," to use Edgeworth's words, we must distinguish between two separate concepts: (i) sympathy and (ii) commitment. The former corresponds to the case in which the concern for others directly affects one's own welfare. If the knowledge of torture of others makes you sick, it is a case of sympathy; if it does not make you feel personally worse off, but you think it is wrong and you are ready to do something to stop it, it is a case of commitment. I do not wish to claim that the words chosen have any very great merit, but the distinction is, I think, important. It can be argued that behavior based on sympathy is in an important sense egoistic, for one is oneself pleased at others' pleasure and pained at others' pain, and the pursuit of one's own utility may thus be helped by sympathetic action. It is action based on commitment rather than sympathy which would be non-egoistic in this sense. (Note, however, that the *existence* of sympathy does not imply that the action helpful to others must be *based on* sympathy in the sense that the action would not take place had one

---

Programme in the Theory of Consumer Behaviour" (Ph.D. thesis, Cambridge University, 1975, forthcoming). See also the pragmatic criticisms of Kornai, *Anti-Equilibrium*, chap. 11.

got less or no comfort from others' welfare. This question of *causation* is to be taken up presently.)

Sympathy is, in some ways, an easier concept to analyze than commitment. When a person's sense of well-being is psychologically dependent on someone else's welfare, it is a case of sympathy; other things given, the awareness of the increase in the welfare of the other person then makes this person directly better off. (Of course, when the influence is negative, the relation is better named "antipathy," but we can economize on terminology and stick to the term "sympathy," just noting that the relation can be positive or negative.) While sympathy relates similar things to each other–namely, welfares of different persons–commitment relates choice to anticipated levels of welfare. One way of defining commitment is in terms of a person choosing an act that he believes will yield a lower level of personal welfare to him than an alternative that is also available to him. Notice that the comparison is between *anticipated* welfare levels, and therefore this definition of commitment excludes acts that go against self-interest resulting purely from a failure to foresee consequences.

A more difficult question arises when a person's choice happens to coincide with the maximization of his anticipated personal welfare, but that is not the *reason* for his choice. If we wish to make room for this, we can expand the definition of commitment to include cases in which the person's choice, while maximizing anticipated personal welfare, would be unaffected under at least one counterfactual condition in which the act chosen would cease to maximize personal welfare. Commitment in this more inclusive sense may be difficult to ascertain not only in the context of others' choices but also in that of one's own, since it is not always clear what one would have done had the circumstances been different. This broader sense may have particular relevance when one acts on the basis of a concern for duty which, if violated, could cause remorse, but the action is really chosen out of the sense of duty rather than just to avoid the illfare resulting from the remorse that would occur if one were to act otherwise. (Of course, even the narrower sense of commitment will cover the case in which the illfare resulting from the remorse, if any, is *outweighed* by the gain in welfare.)

I have not yet referred to uncertainty concerning anticipated wel-

fare. When this is introduced, the concept of sympathy is unaffected, but commitment will require reformulation. The necessary modifications will depend on the person's reaction to uncertainty. The simplest case is probably the one in which the person's idea of what a "lottery" offers to him in terms of personal gain is captured by the "expected utility" of personal welfare (that is, adding personal welfares from different outcomes weighted by the probability of occurrence of each outcome). In this case, the entire discussion is reformulated simply replacing personal welfare by *expected* personal welfare; commitment then involves choosing an action that yields a lower expected welfare than an alternative available action. (The broader sense can also be correspondingly modified.)

In the terminology of modern economic theory, sympathy is a case of "externality." Many models rule out externalities, for example, the standard model to establish that each competitive equilibrium is a Pareto optimum and belongs to the core of the economy. If the existence of sympathy were to be permitted in these models, some of these standard results would be upset, though by no means all of them.[20] But this would not require a serious revision of the basic structure of these models. On the other hand, commitment does involve, in a very real sense, counterpreferential choice, destroying the crucial assumption that a chosen alternative must be better than (or at least as good as) the others for the person choosing it, and this would certainly require that models be formulated in an essentially different way.

The contrast between sympathy and commitment may be illustrated with the story of two boys who find two apples, one large, one small. Boy *A* tells boy *B*, "You choose." *B* immediately picks the larger apple. *A* is upset and permits himself the remark that this was grossly unfair. "Why?" asks *B*. "Which one would *you* have chosen, if you were to choose rather than me?" "The smaller one, of course," *A* replies. *B* is now triumphant: "Then what are you complaining about? That's the one you've got!" *B* certainly wins this round of the argument, but in

20. See A.K. Sen, "Labour Allocation in a Co-operative Enterprise," *Review of Economic Studies* 33 (1966); S.G. Winter, Jr., "A Simple Remark on the Second Optimality Theorem of Welfare Economics," *Journal of Economic Theory* 1 (1969); Collard, "Edgeworth's Propositions"; G.C. Archibald and D. Donaldson, "Non-paternalism and Basic Theorems of Welfare Economics," *Canadian Journal of Economics* 9 (1976).

fact *A* would have lost nothing from *B*'s choice had his own hypothetical choice of the smaller apple been based on sympathy as opposed to commitment. *A*'s anger indicates that this was probably not the case.

Commitment is, of course, closely connected with one's morals. But moral this question is in a very broad sense, covering a variety of influences from religious to political, from the ill-understood to the well-argued. When, in Bernard Shaw's *The Devil's Disciple*, Judith Anderson interprets Richard Dudgeon's willingness to be hanged in place of her husband as arising from sympathy for him or love for her, Richard is adamant in his denial: "What I did last night, I did in cold blood, caring not half so much for your husband, or for you as I do for myself. I had no motive and no interest: all I can tell you is that when it came to the point whether I would take my neck out of the noose and put another man's into it, I could not do it."[21]

The characteristic of commitment with which I am most concerned here is the fact that it drives a wedge between personal choice and personal welfare, and much of traditional economic theory relies on the identity of the two. This identity is sometimes obscured by the ambiguity of the term "preference," since the normal use of the word permits the identification of preference with the concept of being better off, and at the same time it is not quite unnatural to define "preferred" as "chosen." I have no strong views on the "correct" use of the word "preference," and I would be satisfied as long as both uses are not *simultaneously* made, attempting an empirical assertion by virtue of two definitions.[22] The basic link between choice behavior and welfare achievements in the traditional models is severed as soon as commitment is admitted as an ingredient of choice.

V

"Fine," you might say, "but how relevant is all this to the kind of choices with which economists are concerned? Economics does not have much to do with Richard Dudgeon's march to the gallows." I

21. G.B. Shaw, *Three Plays for Puritans* (Harmondsworth, 1966), p. 94.

22. See my "Behaviour and the Concept of Preference," *Economica* 40 (1973); and Shick, "Rationality and Sociality."

think one should immediately agree that for many types of behavior, commitment is unlikely to be an important ingredient. In the private purchase of many consumer goods, the scope for the exercise of commitment may indeed be limited and may show up rather rarely in such exotic acts as the boycotting of South African avocados or the eschewing of Spanish holidays. Therefore, for many studies of consumer behavior and interpretations thereof, commitment may pose no great problem. Even sympathy may not be extremely important, the sources of interpersonal interdependence lying elsewhere, for example, in the desire to keep up with the Joneses or in being influenced by other people's habits.[23]

But economics is not concerned only with consumer behavior; nor is consumption confined to "private goods." One area in which the question of commitment is most important is that of the so-called public goods. These have to be contrasted with "private goods" which have the characteristic that they cannot be used by more than one person: if you ate a piece of apple pie, I wouldn't consider devouring it too. Not so with "public goods," for example, a road or a public park, which you and I may both be able to use. In many economic models private goods are the only ones around, and this is typically the case when the "invisible hand" is given the task of doing visible good. But, in fact, public goods are important in most economies and cover a wide range of services from roads and street lighting to defense. There is much evidence that the share of public goods in national consumption has grown rather dramatically in most countries in the world.

The problem of optimal allocation of public goods has also been much discussed, especially in the recent economic literature.[24] A lot

23. See J.S. Duesenberry, *Income, Saving and the Theory of Consumer Behavior* (Cambridge, Mass., 1949); S.J. Prais and H.S. Houthakker, *The Analysis of Family Budgets* (Cambridge, 1955); W. Gaertner, "A Dynamic Model of Interdependent Consumer Behaviour," mimeographed (Bielefeld University, 1973); R.A. Pollak, "Interdependent Preferences," *American Economic Review* 66 (1976).

24. See E. Lindahl, *Die Gerechtigkeit der Besteuerung* (Lund, 1919), translated in R.A. Musgrave and A. Peacock, *Classics in the Theory of Public Finance* (London, 1967); P.A. Samuelson, "The Pure Theory of Public Expenditure," *Review of Economic Studies* 21 (1954); R. Musgrave, *The Theory of Public Finance* (New York, 1959); L. Johansen, *Public Economics* (Amsterdam, 1966); D.K. Foley, "Lindahl's Solution and the Core of an Economy with Public Goods,"

of attention, in particular, has been devoted to the problem of correct revelation of preferences. This arises most obviously in the case of subscription schemes where a person is charged according to benefits received. The main problem centers on the fact that it is in everybody's interest to understate the benefit he expects, but this understatement may lead to the rejection of a public project which would have been justified if true benefits were known. Analysis of this difficulty, sometimes referred to as the "free rider" problem, has recently led to some extremely ingenious proposals for circumventing this inefficiency within the framework of egoistic action.[25] The reward mechanism is set up with such ungodly cunning that people have an incentive to reveal exactly their true willingness to pay for the public good in question. One difficulty in this solution arises from an assumed limitation of strategic possibilities open to the individual, the removal of which leads to an impossibility result.[26] Another difficulty concerns the fact that in giving people the incentive to reveal the truth, money is handed out and the income distribution shifts in a way unguided by distributional considerations. This effect can, of course, be undone by a redistribution of initial endowments and profit shares,[27] but that action obviously raises difficulties of its own.

Central to this problem is the assumption that when asked a ques-

Econometrica, 38 (1970); E. Malinvaud, "Prices for Individual Consumption, Quantity Indicators for Collective Consumption," *Review of Economic Studies* 39 (1972).

25. T. Groves and J. Ledyard, "Optimal Allocation of Public Goods: A Solution to the 'Free Rider Problem,'" Discussion Paper No. 144 (Center for Mathematical Studies in Economics and Management Science, Northwestern University, 1975); J. Green and J.J. Laffont, "On the Revelation of Preference for Public Goods," Technical Report No. 140 (Institute for Mathematical Studies in the Social Sciences, Stanford University, 1974). See also J. Dreze and D. de la Vallee Poussin, "A Tatonnement Process for Public Goods," *Review of Economic Studies* 38 (1971); E. Malinvaud, "A Planning Approach to the Public Goods Problem," *Swedish Journal of Economics* 73 (1971); V.L. Smith, "Incentive Compatible Experimental Processes for the Provision of Public Goods," mimeographed (Econometric Society Summer Meeting, Madison, 1976).

26. See J. Ledyard and D.J. Roberts, "On the Incentive Problem for Public Goods," Discussion Paper No. 116 (CMSEMS, Northwestern University, 1974). See also L. Hurwicz, "On Informationally Decentralized Systems," in R. Radner and B. McGuire, *Decisions and Organizations* (Amsterdam, 1972).

27. See Theorem 4.2 in Groves and Ledyard, "Optimal Allocation of Public Goods."

tion, the individual gives that answer which will maximize his personal gain. How good is this assumption? I doubt that in general it is very good. ("Where is the railway station?" he asks me. "There," I say, pointing at the post office, "and would you please post this letter for me on the way?" "Yes," he says, determined to open the envelope and check whether it contains something valuable.) Even in the particular context of revelation of preferences for public goods the gains-maximizing behavior may not be the best assumption. Leif Johansen, one of the major contributors to public economics, is, I think, right to question the assumption in this context:

> Economic theory in this, as well as in some other fields, tends to suggest that people are honest only to the extent that they have economic incentives for being so. This is a homo oeconomicus assumption which is far from being obviously true, and which needs confrontation with observed realities. In fact, a simple line of thought suggests that the assumption can hardly be true in its most extreme form. No society would be viable without some norms and rules of conduct. Such norms and rules are necessary for viability exactly in fields where strictly economic incentives are absent and cannot be created.[28]

What is at issue is not whether people invariably give an honest answer to every question, but whether they always give a gains-maximizing answer, or at any rate, whether they give gains-maximizing answers often enough to make that the appropriate general assumption for economic theory. The presence of non-gains-maximizing answers, including truthful ones, immediately brings in commitment as a part of behavior.

The question is relevant also to the recent literature on strategic voting. A number of beautiful analytical results have recently been established showing the impossibility of any voting procedure satisfying certain elementary requirements and making honest voting the

28. L. Johansen, "The Theory of Public Goods: Misplaced Emphasis" (Institute of Economics, University of Oslo, 1976). See also J.J. Laffont, "Macroeconomic Constraints, Economic Efficiency and Ethics," mimeographed (Harvard University, 1974); P. Bohm, "Estimating Demand for Public Goods: An Experiment," *European Economic Review* 3 (1972).

gains-maximizing strategy for everyone.[29] The correctness of these results is not in dispute, but is it appropriate to assume that people always do try to maximize personal gains in their voting behavior? Indeed, in large elections, it is difficult to show that any voter has any real prospect of affecting the outcome by his vote, and if voting involves some cost, the expected net gain from voting may typically be negative. Nevertheless, the proportion of turnout in large elections may still be quite high, and I have tried to argue elsewhere that in such elections people may often be "guided not so much by maximization of expected utility, but something much simpler, viz, just a desire to record one's true preference."[30] If this desire reflects a sense of commitment, then the behavior in question would be at variance with the view of man in traditional economic theory.

VI

The question of commitment is important in a number of other economic contexts.[31] It is central to the problem of work motivation, the importance of which for production performance can hardly be ignored.

It is certainly costly and may be impossible to devise a system of

29. A. Gibbard, "Manipulation of Voting Schemes: A General Result," *Econometrica* 41 (1973); M.A. Satterthwaite, "Strategy-proofness and Arrow's Conditions," *Journal of Economic Theory* 10 (1975); D. Schmeidler and H. Sonnenschein, "The Possibility of Non-manipulable Social Choice Functions" (CMSEMS, Northwestern University, 1974); B. Dutta and P.K. Pattanaik, "On Nicely Consistent Voting Systems" (Delhi School of Economics, 1975); P.K. Pattanaik, "Strategic Voting without Collusion under Binary and Democratic Group Decision Rules," *Review of Economic Studies* 42 (1975); B. Peleg, "Consistent Voting Systems" (Institute of Mathematics, Hebrew University, Jerusalem, 1976); A. Gibbard, "Social Decision, Strategic Behavior, and Best Outcomes: An Impossibility Result," Discussion Paper No. 224 (CMSEMS, Northwestern University, 1976).

30. See A.K. Sen, *Collective Choice and Social Welfare* (Edinburgh and San Francisco, 1970), p. 195.

31. See Ragnar Frisch's discussion of the need for "a realistic theoretical foundation for social policy" in his "Samarbeid mellom Politikere og Økonometrikere om Formuleringen av Politiske Preferenenser" (Socialøkonomen, 1971). (I am grateful to Leif Johansen for translating the relevant portions of the paper for me.) See also J.A. Mirrlees, "The Economics of Charitable Contributions," Econometric Society European meeting (Oslo, 1973).

supervision with rewards and punishment such that everyone has the incentive to exert himself. Every economic system has, therefore, tended to rely on the existence of attitudes toward work which supersedes the calculation of net gain from each unit of exertion. Social conditioning plays an extremely important part here.[32] I am persuaded that Britain's present economic difficulties have a great deal to do with work-motivation problems that lie outside the economics of rewards and punishments, and one reason why economists seem to have so little to contribute in this area is the neglect in traditional economic theory of this whole issue of commitment and the social relations surrounding it.[33]

These questions are connected, of course, with ethics, since moral reasoning influences one's actions, but in a broader sense these are matters of culture, of which morality is one part. Indeed, to take an extreme case, in the Chinese "cultural revolution" one of the primary aims was the increase of the sense of commitment with an eye on economic results: "the aim of the Great Proletarian Cultural Revolution is to revolutionize people's ideology and as a consequence to achieve greater, faster, better and more economical results in all fields of work."[34] Of course, China was experimenting with reducing dramatically the role of material incentives in production, which would

32. See A. Fox, *Beyond Contract: Work, Power and Trust Relations* (London, 1974); H.G. Nutzinger, "The Firm as a Social Institution: The Failure of a Contractarian Viewpoint," Working Paper No. 52 (Alfred Weber Institute, University of Heidelberg, 1976).

33. Cf. "Nor . . . should we forget the extent to which conventional theory ignores how and why work is organized within the firm and establishment in the way it is, what may be called the 'social relations' of the production process," R.A. Gordon, "Rigor and Relevance in a Changing Institutional Setting," Presidential Address, *American Economic Review* 66 (1976). See also R. Dahrendorf, *Class and Class Conflict in Industrial Society* (Stanford, 1959); O.E. Williamson, "The Evolution of Hierarchy: An Essay on the Organization of Work," Fels Discussion Paper No. 91 (University of Pennsylvania, 1976); and S.A. Marglin, "What Do Bosses Do? The Origins and Functions of Hierarchy in Capitalist Production," *Review of Radical Political Economics* 6 (1974).

34. "The Decision of the Central Committee of the Chinese Communist Party Concerning the Great Proletarian Cultural Revolution," adopted on 8 August 1966, reproduced in Joan Robinson, *The Cultural Revolution in China* (Harmondsworth, 1969). See also A.K. Sen, *On Economic Inequality* (Oxford, 1973); and C. Riskin, "Maoism and Motivation: A Discussion of Work Motivation in China," *Bulletin of Concerned Asian Scholars*, 1973.

certainly have increased the part that commitment was meant to play, but even within the traditional systems of payments, much reliance is usually placed on rules of conduct and modes of behavior that go beyond strictly economic incentives.[35] To run an organization *entirely* on incentives to personal gain is pretty much a hopeless task.

I will have a bit more to say presently on what might lie behind the sense of commitment, but I would like to emphasize at this stage that the morality or culture underlying it may well be of a limited kind—far removed from the grandeur of approaches such as utilitarianism. The "implicit collusions" that have been observed in business behavior in oligopolies seem to work on the basis of a system of mutual trust and sense of responsibility which has well-defined limits, and attempts at "universalization" of the same kind of behavior in other spheres of action may not go with it at all. There it is strictly a question of business ethics which is taken to apply within a fairly limited domain.

Similarly, in wage negotiations and in collective bargaining the sense of solidarity on either side may have well-defined limits, and may not fit in at all with an approach such as that of general utilitarianism. Edgeworth's implicit assumption, on which I commented earlier, that egoism and utilitarianism exhaust the possible alternative motivations, will be especially unhelpful in this context. While the field of commitment may be large, that of commitment based on utilitarianism and other universalized moral systems may well form a relatively small part of it.

VII

The economic theory of utility, which relates to the theory of rational behavior, is sometimes criticized for having too much structure; human beings are alleged to be "simpler" in reality. If our argument so far has been correct, precisely the opposite seems to be the case: traditional theory has *too little* structure. A person is given *one* preference ordering, and as and when the need arises this is supposed to reflect his interests, represent his welfare, summarize his idea of what should be done, and describe his actual choices and behavior. Can one prefer-

35. See Williamson, "The Evolution of Hierarchy," for a critical analysis of the recent literature in this area.

ence ordering do all these things? A person thus described may be "rational" in the limited sense of revealing no inconsistencies in his choice behavior, but if he has no use for these distinctions between quite different concepts, he must be a bit of a fool. The *purely* economic man is indeed close to being a social moron. Economic theory has been much preoccupied with this rational fool decked in the glory of his *one* all-purpose preference ordering. To make room for the different concepts related to his behavior we need a more elaborate structure.

What kind of a structure do we need? A bit more room up top is provided by John Harsanyi's important distinction between a person's "ethical" preferences and his "subjective" preferences: "the former must express what this individual prefers (or, rather would prefer), on the basis of impersonal social considerations alone, and the latter must express what he actually prefers, whether on the basis of his personal interests or on any other basis."[36] This dual structure permits us to distinguish between what a person thinks is good from the social point of view and what he regards as good from his own personal point of view. Presumably sympathy enters directly into the so-called subjective preference, but the role of commitment is left somewhat unclear. Insofar as a person's "subjective" preferences are taken to "define his utility function," the intention seems to be to exclude commitment from it, but an ambiguity arises from the fact that these are defined to "express his preferences in the full sense of the word as they actually are." Is this in the sense of choice, or in the sense of his conception of his own welfare? Perhaps Harsanyi intended the latter, since "ethical" preferences are by contrast given the role of expressing "what he prefers only in those possibly rare moments when he forces a special impartial and impersonal attitude on himself."[37] But what if he departs from his personal welfare maximization (including any sympathy), not through an impartial concern for all,[38] but through a sense of commitment to

36. J. Harsanyi, "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *Journal of Political Economy* 63 (1955): 315.

37. Ibid., pp. 315-316.

38. Note that for Harsanyi "an individual's preferences satisfy this requirement of impersonality if they indicate what social situation he would choose if he did not know what his general position would be in the new situation chosen

some particular group, say to the neighborhood or to the social class to which he belongs? The fact is we are still short of structure.

Even in expressing moral judgments from an impersonal point of view, a *dual* structure is deficient. Surely a preference ordering can be *more* ethical than another but *less* so than a third. We need more structure in this respect also. I have proposed elsewhere—at the 1972 Bristol conference on "practical reason"—that we need to consider *rankings of preference rankings* to express our moral judgments.[39] I would like to discuss this structure a bit more. A particular morality can be viewed, not just in terms of the "most moral" ranking of the set of alternative actions, but as a moral ranking of the rankings of actions (going well beyond the identification merely of the "most moral" ranking of actions). Let X be the set of alternative and mutually exclusive combinations of actions under consideration, and let Y be the set of rankings of the elements of X. A ranking of the set Y (consisting of action-rankings) will be called a meta-ranking of action-set X. It is my claim that a particular ranking of the action-set X is not articulate enough to express much about a given morality, and a more robust format is provided by choosing a meta-ranking of actions (that is, a ranking of Y rather than of X). Of course, such a meta-ranking may include *inter alia* the specification of a particular action-ranking as the "most moral," but insofar as actual behavior may be based on a compromise between claims of morality and the pursuit of various other objectives (including self-interest), one has to look also at the relative moral standings of those action-rankings that are *not* "most moral."

To illustrate, consider a set X of alternative action combinations and the following three rankings of this action-set X: ranking A represent-

---

(and in any of its alternatives) but rather had an equal chance of obtaining any of the social positions existing in this situation, from the highest down to the lowest" (p. 316).

39. A.K. Sen, "Choice, Orderings and Morality," in S. Körner, ed., *Practical Reason* (Oxford, 1974). See also J. Watkins' rejoinder and my reply in the same volume, and R.C. Jeffrey, "Preferences among Preferences," *Journal of Philosophy* 71 (1974); K. Binmore, "An Example in Group Preference," *Journal of Economic Theory* 10 (1975); and B.A. Weisbrod, "Toward a State-Preference Model of Utility Function Preferences: A Conceptual Note," mimeographed (University of Wisconsin, 1976).

ing my personal welfare ordering (thus, in some sense, representing my personal interests), ranking B reflecting my "isolated" personal interests ignoring sympathy (when such a separation is possible, which is not always so),[40] and ranking C in terms of which actual choices are made by me (when such choices are representable by a ranking, which again is not always so).[41] The "most moral" ranking M can, conceivably, be any of these rankings A, B, or C. Or else it can be some other ranking quite distinct from all three. (This will be the case if the actual choices of actions are not the "most moral" in terms of the moral system in question, and if, furthermore, the moral system requires sacrifice of some self-interest and also of "isolated" self-interest.) But even when some ranking M distinct from A, B, and C is identified as being at the top of the moral table, that still leaves open the question as to how A, B, and C may be ordered vis-à-vis each other. If, to take a particular example, it so happens that the pursuit of self-interest, including pleasure and pain from sympathy, is put morally above the pursuit of "isolated" self-interest (thereby leading to a partial coincidence of self-interest with morality), and the actual choices reflect a morally superior position to the pursuit of self-interest (perhaps due to a compromise in the moral direction), then the morality in question precipitates the meta-ranking M, C, A, B, in descending order. This, of course, goes well beyond specifying that M is "morally best."

The technique of meta-ranking permits a varying extent of moral articulation. It is not being claimed that a moral meta-ranking must be a *complete* ordering of the set Y, that is, must completely order all rankings of X. It can be a *partial* ordering, and I expect it often will be incomplete, but I should think that in most cases there will be no problem in going well beyond the limited expression permitted by the twofold specification of "ethical" and "subjective" preferences.

The rankings of action can, of course, be ordered also on grounds other than a particular system of morality: meta-ranking is a general

40. This presupposes some "independence" among the different elements influencing the level of overall welfare, implying some "separability." See W.M. Gorman, "Tricks with Utility Functions," in M. Artis and A.R. Nobay, eds., *Essays in Economic Analysis* (Cambridge, 1975).

41. See fn. 16 above.

technique usable under alternative interpretations of the meta-ranking relation. It can be used to describe a particular ideology or a set of political priorities or a system of class interests. In quite a different context, it can provide the format for expressing what preferences one would have preferred to have ("I wish I liked vegetarian foods more," or "I wish I didn't enjoy smoking so much"). Or it can be used to analyze the conflicts involved in addiction ("Given my current tastes, I am better off with heroin, but having heroin leads me to addiction, and I would have preferred not to have these tastes"). The tool of meta-rankings can be used in many different ways in distinct contexts.

This is clearly not the occasion to go into a detailed analysis of how this broader structure permits a better understanding of preference and behavior. A structure is not, of course, a theory, and alternative theories can be formulated using this structure. I should mention, however, that the structure demands much more information than is yielded by the observation of people's actual choices, which would at most reveal only the ranking C. It gives a role to introspection and to communication. To illustrate one use of the apparatus, I may refer to some technical results. Suppose I am trying to investigate your conception of your own welfare. You first specify the ranking A which represents your welfare ordering. But I want to go further and get an idea of your *cardinal* utility function, that is, roughly speaking, not only which ranking gives you more welfare but also by how much. I now ask you to order the different rankings in terms of their "closeness" to your actual welfare ranking A, much as a policeman uses the technique of photofit: is this more like him, or is that? If your answers reflect the fact that reversing a stronger preference makes the result more distant than reversing a weaker intensity of preference, your replies will satisfy certain consistency properties, and the order of rankings will permit us to compare your welfare *differences* between pairs. In fact, by considering higher and higher order rankings, we can determine your cardinal welfare function as closely as you care to specify.[42] I am not saying that this type of dialogue is the best way of discovering your welfare function, but it does illustrate that once we give

42. This result and some related ones emerged in discussions with Ken Binmore in 1975, but a projected joint paper reporting them is still, alas, unwritten. More work on this is currently being done also by R. Nader-Ispahani.

up the assumption that observing choices is the only source of data on welfare, a whole new world opens up, liberating us from the informational shackles of the traditional approach.

This broader structure has many other uses, for example, permitting a clearer analysis of *akrasia*—the weakness of will—and clarifying some conflicting considerations in the theory of liberty, which I have tried to discuss elsewhere.[43] It also helps in analyzing the development of behavior involving commitment in situations characterized by games such as the Prisoners' Dilemma.[44] This game is often treated, with some justice, as the classic case of failure of individualistic rationality. There are two players and each has two strategies, which we may call selfish and unselfish to make it easy to remember without my having to go into too much detail. Each player is better off personally by playing the selfish strategy *no* matter what the other does, but both are better off if both choose the unselfish rather than the selfish strategy. It is individually optimal to do the selfish thing: one can only affect one's own action and not that of the other, and given the other's strategy—no matter what—each player is better off being selfish. But this combination of selfish strategies, which results from self-seeking by both, produces an outcome that is worse for both than the result of both choosing the unselfish strategy. It can be shown that this conflict can exist even if the game is repeated many times.

Some people find it puzzling that individual self-seeking by each should produce an inferior outcome for all, but this, of course, is a well-known conflict, and has been discussed in general terms for a very long time. Indeed, it was the basis of Rousseau's famous distinc-

43. See Sen, "Choice, Orderings and Morality"; and also Sen, "Liberty, Unanimity and Rights," *Economica* 43 (1976). Note also the relevance of this structure in analyzing the incompleteness of the conception of liberty in terms of the ability to do what one *actually wishes*. Cf. "If I find that I am able to do little or nothing of what I wish, I need only contract or extinguish my wishes, and I am made free. If the tyrant (or 'hidden persuader') manages to condition his subjects (or customers) into losing their original wishes and embrace ('internalize') the form of life he has invented for them, he will, on this definition, have succeeded in liberating them." I. Berlin, "Two Concepts of Liberty," in *Four Essays on Liberty* (Oxford, 1969), pp. 139–140).

44. See R.D. Luce and H. Raiffa, *Games and Decisions* (New York, 1958); A. Rapoport and A.M. Chammah, *Prisoner's Dilemma: A Study in Conflict and Cooperation* (Ann Arbor, 1965); W.G. Runciman and A.K. Sen, "Games, Justice and the General Will," *Mind*, 74 (1965); N. Howard, *Paradoxes of Rationality* (Cambridge, Mass., 1971).

tion between the "general will" and the "will of all."[45] But the puzzle from the point of view of rational behavior lies in the fact that in actual situations people often do not follow the selfish strategy. Real life examples of this type of behavior in complex circumstances are well known, but even in controlled experiments in laboratory conditions people playing the Prisoners' Dilemma frequently do the unselfish thing.[46]

In interpreting these experimental results, the game theorist is tempted to put it down to the lack of intelligence of the players: "Evidently the run-of-the-mill players are not strategically sophisticated enough to have figured out that strategy DD [the selfish strategy] is the only rationally defensible strategy, and this intellectual short-coming saves them from losing."[47] A more fruitful approach may lie in permitting the possibility that the person is *more* sophisticated than the theory allows and that he has asked himself what type of preference he would like the other player to have, and on somewhat Kantian grounds has considered the case for himself having those preferences, or behaving *as if* he had them. This line of reasoning requires him to consider the modifications of the game that would be brought about by acting through commitment (in terms of "revealed preferences," this would look *as if* he had different preferences from the ones he actually had), and he has to assess alternative behavior norms in that light. I have discussed these issues elsewhere;[48] thus I shall simply note here that the apparatus of *ranking of rankings* assists the reasoning which involves considering the merits of having different types of preferences (or of acting as if one had them).

VIII

Admitting behavior based on commitment would, of course have far-reaching consequences on the nature of many economic models. I

45. See Runciman and Sen.

46. See, for example, L.B. Lave, "An Empirical Approach to the Prisoner's Dilemma Game," *Quarterly Journal of Economics* 76 (1962), and Rapoport and Chammah, *Prisoner's Dilemma.*

47. Rapoport and Chammah, p. 29.

48. Sen, "Choice, Orderings and Morality." See also K. Baier, "Rationality and Morality," and A.K. Sen, "Rationality and Morality: A Reply," both forthcoming in *Erkenntnis*; K. Baier, *The Moral Point of View* (Ithaca, 1958); and Fred Schick's analysis, "Rationality and Sociality."

have tried to show why this change is necessary and why the consequences may well be serious. Many issues remain unresolved, including the empirical importance of commitment as a part of behavior, which would vary, as I have argued, from field to field. I have also indicated why the empirical evidence for this cannot be sought in the mere observation of actual choices, and must involve other sources of information, including introspection and discussion.

There remains, however, the issue as to whether this view of man amounts to seeing him as an irrational creature. Much depends on the concept of rationality used, and many alternative characterizations exist. In the sense of *consistency* of choice, there is no reason to think that admitting commitment must imply any departure from rationality. This is, however, a weak sense of rationality.

The other concept of rationality prevalent in economics identifies it with the possibility of justifying each act in terms of self-interest: when act x is chosen by person i and act y rejected, this implies that i's personal interests are expected by i to be better served by x than by y. There are, it seems to me, three distinct elements in this approach. First, it is a consequentialist view: judging acts by conseqences only.[49] Second, it is an approach of *act* evaluation rather than *rule* evaluation. And third, the only consequences considered in evaluating acts are those on one's own interests, everything else being at best an intermediate product. It is clearly possible to dispute the claims of each of these elements to being a necessary part of the conception of rationality in the dictionary sense of "the power of being able to exercise one's reason." Moreover, arguments for rejecting the straightjacket of each of these three principles are not hard to find. The case for actions based on commitment can arise from the violation of any of these three principles. Commitment sometimes relates to a sense of obligation going beyond the consequences. Sometimes the lack of personal gain in particular *acts* is accepted by considering the value of *rules* of behavior. But even within a consequentialist act-evaluation framework, the exclusion of any consideration other than self-interest seems to impose a wholly arbitrary limitation on the notion of rationality.

49. On the nature of "consequentialism" and problems engendered by it, see B. Williams, "A Critique of Utilitarianism," in J.J.C. Smart and B. Williams, *Utilitarianism: For and Against* (Cambridge, 1973).

Henry Sidgwick noted the arbitrary nature of the assumption of egoism:

> If the Utilitarian has to answer the question, "Why should I sacrifice my own happiness for the greater happiness of another?" it must surely be admissible to ask the Egoist, "Why should I sacrifice a present pleasure for one in the future? Why should I concern myself about my own future feelings any more than about the feelings of other persons?" It undoubtedly seems to Common Sense paradoxical to ask for a reason why one should seek one's own happiness on the whole; but I do not see how the demand can be repudiated as absurd by those who adopt views of the extreme empirical school of psychologists, although those views are commonly supposed to have a close affinity with Egoistic Hedonism. Grant that the Ego is merely a system of coherent phenomena, that the permanent identical "I" is not a fact but a fiction, as Hume and his followers maintain; why, then, should one part of the series of feelings into which the Ego is resolved be concerned with another part of the same series, any more than with any other series?[50]

The view of rationality that identifies it with consequentialist act-evaluation using self-interest can be questioned from any of these three angles. Admitting commitment as a part of behavior implies no denial of reasoned assessment as a basis for action.

There is not much merit in spending a lot of effort in debating the "proper" definition of rationality. The term is used in many different senses, and none of the criticisms of the behavioral foundations of economic theory presented here stands or falls on the definition chosen. The main issue is the acceptability of the assumption of the invariable pursuit of self-interest in each act. Calling that type of behavior rational, or departures from it irrational, does not change the relevance of these criticisms, though it does produce an arbitrarily narrow definition of rationality. This paper has not been concerned with

50. H. Sidgwick, *The Method of Ethics* (London, 1874; 7th ed., 1907), pp. 418-419. See also Nagel's forceful exposition of the thesis that "altruism itself depends on a recogniton of the reality of other persons, and on the equivalent capacity to regard oneself as merely one individual among many." *The Possibility of Altruism*, p. 1.

the question as to whether human behavior is better described as rational or irrational. The main thesis has been the need to accommodate commitment as a part of behavior. Commitment does not presuppose reasoning, but it does not exclude it; in fact, insofar as consequences on others have to be more clearly understood and assessed in terms of one's values and instincts, the scope for reasoning may well expand. I have tried to analyze the structural extensions in the conception of preference made necessary by behavior based on reasoned assessment of commitment. Preferences as rankings have to be replaced by a richer structure involving meta-rankings and related concepts.

I have also argued against viewing behavior in terms of the traditional dichotomy between egoism and universalized moral systems (such as utilitarianism). Groups intermediate between oneself and all, such as class and community, provide the focus of many actions involving commitment. The rejection of egoism as description of motivation does not, therefore, imply the acceptance of some universalized morality as the basis of actual behavior. Nor does it make human beings excessively noble.

Nor, of course, does the use of reasoning imply remarkable wisdom.

> It is as true as Caesar's name was Kaiser,
> That no economist was ever wiser,

said Robert Frost in playful praise of the contemporary economist. Perhaps a similarly dubious tribute can be paid to the economic man in our modified conception. If he shines at all, he shines in comparison –in contrast–with the dominant image of the rational fool.