Public Choice 51: 3-23 (1986).

Presidential Address

Rational egoism versus adaptive egoism as fundamental postulate for a descriptive theory of human behavior

DENNIS C. MUELLER*
Department of Economics, University of Maryland, College Park, MD 20742

The assumption that individuals, be they voters, politicians or bureaucrats, act rationally in their own self-interest is the most important and obvious characteristic of public choice distinguishing it from its sister disciplines in the social sciences. By and large the assumption has served us well, allowing us to derive numerous propositions which are neither trivial nor obvious. In recent years, we have reinforced the plausibility of our analytic models with empirical evidence obtained both from econometric analyses of secondary data, and from primary data manufactured in the experimental laboratory. During the last generation the harvest of analytical, empirical and experimental results has allowed public choice to develop as a sort of *Wunderkind* within the family of social sciences, often receiving both the praise and resentful envy a *Wunderkind* inevitably receives.

Although public choice's rise as a scientific subdiscipline has been impressive, its record is not without blemish. Four years ago in his presidential address to the Public Choice Society John Ledyard set out to slay two dragons that have stalked public choice since its birth: that rational voters do not vote, and that when rational candidates compete for the votes of rational voters, there is no equilibrium outcome (1984). The fruit of his search was a proof that under certain, reasonable assumptions competition of candidates for votes does lead to an equilibrium outcome. At this equilibrium, however, no voter votes. The first result and similar results by my colleague, Peter Coughlin (1982, with Nitzan, 1981), I consider among the most important and exciting theorems in public choice to appear in the last few years. They constitute, I believe, an important component of the answer to Gordon Tullock's question, 'Why so much stability?' (1981), which is to say that these theorems have empirical relevance as well as theoretical elegance. The second result, as an empirical prediction, is somewhat embarrassing. In

* Revised version of the Presidential Address presented to the Public Choice Society, Baltimore, March 22, 1986. Support for the research on this paper was received in part from the Thyssen Foundation of Germany and the Sloan Foundation Workshop grant to the Department of Economics, University of Maryland, and the Thyssen Foundation of Germany. Helpful comments were received from Jack Hirschleifer, Norbert Hornstein, Hartmut Kliemt, Viktor Vanberg, and Ulrich Witt.

a presidential election it is off by some 80 million or so.

There are other examples of empirical predictions stemming from the rational egoist assumption, which are equally wide of the mark, as, for example, the large voluntary contributions to the provision of public goods, which have been observed in experimental situations. What is more, we could easily add to the list. Why do people bother to queue to buy tickets or to get on a bus, when they could obtain a better seat by barging to the front? Why do people stop for a stop sign, when no policeman is in sight, and the intersection is empty? Now, of course, some people do barge forward in queues, and some run stop signs. Thus, the behavior of some people is consistent with what the rational egoist assumption predicts. The question the scientist must address when employing the rational egoist postulate is why everyone does not barge to the front? That a few do and generally succeed in getting in ahead of those who do not only adds to the mystery of why the others remain in an orderly queue.

What I suggest in this paper is an extension of the framework used to analyze individual behavior, which would allow us to explain and predict decisions like voting, contributing to public goods, and queuing in a more satisfactory manner than we now can.

1. The prisoners' dilemma

The above examples are types of cooperative behavior that can be characterized by the prisoners' dilemma, and it is common to analyze the behavior of rational egoists in these situations using this analytical tool. As we all know, the noncooperative outcome, not voting, not contributing to the public good, not waiting in the queue, dominates for a single play of the game. Cooperative behavior is usually rationalized by assuming an unbounded or indeterminate number of plays of the game, and assuming that each player adopts a supergame strategy of, say, matching the strategy the others played in the previous round of the game. There are two conceptual difficulties to accepting this explanation for why rational-egoistic individuals behave cooperatively. First, although few prisoners' dilemma games are played in real life for a fixed number of trials, a rational individual of 75 should realize that the expected number of games left to be played is significantly less than for that of an individual of 25. The prisoners' dilemma supergame rationale for cooperative behavior ought to lead to the prediction that the elderly are the most flagrant violators of our social mores, barging ahead in every queue, shoplifting at every chance. Even adding other obvious and not so obvious variables to the equation would not, I suspect, turn this prediction into an empirically verifiable proposition.

Second, and more importantly, the prisoners' dilemma-supergame rationale for cooperative behavior unravels rather quickly as the number of players in the game increases. With large numbers, one player's defection

has an imperceptible impact on the outcome of the game, and should not induce defections by other players. Thus, all rational players should defect.[1]

2. The behaviorist psychology alternative

Most economists, and I presume by extension public choice scholars, seem to aspire to an elegancy in model building and accuracy in prediction they believe to be characteristic of physics. Perhaps, my message at this juncture will carry more weight then if I begin by quoting the advice of a physicist given to social scientists. Percy Bridgeman suggested that 'the principal problem in understanding the actions of men is to understand how they think – how their minds work.'[2]

The prisoners' dilemma breaks down as a predictor of individual behavior, because it is based on an inaccurate description of how the minds of individuals work in prisoners' dilemma situations. Now the criticism, 'people don't really make decisions that way' as a reaction to a rigorous modeling of rational egoist behavior is one familiar to all economists. The usual response to this criticism is that it is not important that the manager consciously equates marginal cost and marginal revenue, but that he behaves *as if* he did. The true test of the model is if its predictions match reality, not its assumptions. Does price rise when a tariff is imposed? But, of course, we are concerned with the prisoners' dilemma model here precisely because its predictions do not receive overwhelming empirical support. Pretending individuals make decisions as if they were rational egoists playing a prisoners' dilemma game may be the root of the empirical problem.[3]

The prisoners' dilemma is a popular analytic tool for analyzing individual behavior because it seems to fit so many everyday situations we confront. Most of us choose the cooperative strategy most of the time. Why? Because we were taught to do so. Our first contact with prisoners' dilemma situations occurred as children. We were taught to pick up our clothes, be quiet, not throw food, etc., by being rewarded when we did so and punished when we did not. Cooperative behavior was thus reinforced, noncooperative behavior punished. Preferring rewards to punishment, we increased the frequency of cooperative behavior, and reduced the frequency of noncooperative behavior.

The reader has probably noted that I have snuck in some of the jargon of behaviorist psychology. A part of what I want to suggest is that we in public choice should rely more on behaviorist psychology to explain and predict individual behavior in prisoners' dilemma situations, and less on game theory.

Behaviorist psychology teaches us that individual acts which are followed by positive reinforcers (rewards) increase in frequency, those followed by punishment decline in frequency. Through this type of operant conditioning we first learn to behave cooperatively in various prisoners' dilemma contexts. Once a mode of behavior has been conditioned, it can be maintained

at high levels of frequency without being rewarded each time it is performed. The child need not be read a story each time it picks up its room. Individuals learn to *generalize*. The child who is rewarded for helping carry the groceries in volunteers to help put the garden tools away. That which is not initially a reward can be *conditioned* to be a reward by frequent association with a reward. The child initially rewarded with a smile and a cookie, eventually behaves cooperatively just to receive a smile.[4]

These principles of psychology should be familiar to every parent and teacher, to anyone who has ever trained a dog. They also help explain the ubiquitous performance of cooperative behavior by individuals in prisoners' dilemma-type situations, in which the noncooperative strategy would appear optimal from a strictly rational egoist perspective. We learn not to steal, to line up and wait our turn, to follow the rules and do what is expected of us as children in the home, in school, and in church. We learn by example, by being told what to do, by being rewarded when we do what is right, punished when we do what is wrong. We learn to emit the cooperative mode of behavior in a variety of situations. As we mature, we are continually confronted with new prisoners' dilemma situations. By generalizing from past prisoners' dilemma situations, we recognize stimuli which often induce us to behave cooperatively. So long as cooperative behavior is sometimes rewarded, and noncooperative behavior sometimes punished, our conditioned responses to being in prisoners' dilemma situations continue to be to emit the cooperative modes of behavior at frequencies which exceed those that occur for those individuals who have not undergone prior operant conditioning.

Except for the choice of words, all of the above is, I am sure, fairly obvious. One is almost embarrassed to make these observations were it not that so many of us who work with rational egoist models continually build our models on assumptions that ignore these truisms from psychology and everyday life. What accounts for our reluctance to make assumptions about individual behavior which allow for conditioned behavior patterns? I think there are at least two explanations. First, we suffer from what those who study innovative activity call 'the not-invented-here bias.' Any hypothesis not developed from within the rational egoism paradigm is viewed with suspicion. Second, even if we give some credence to these alien hypotheses, we fear that to add them to our analytic models would detract from their rigor, make them more difficult to analyze, might even lead to that most brutal and humiliating of all criticisms, the criticism that our models are *ad hoc*. Let me try to some extent to deal with both criticisms. I begin with the question of how to model learned cooperative behavior.

3. On modeling learned cooperative behavior

The usual way to operationalize rational egoism is to assume that an individual maximizes his/her utility subject to some constraints. Consider replacing this assumption in situations in which there are n individuals in a prisoners' dilemma, with the assumption that each individual maximizes an objective function, which is a weighted sum of his/her utility and the utilities of the other $n-1$ individuals in the group. That is, each maximizes

$$O_i = U_i + \theta \sum_{j \neq i}^{n} U_j \tag{1}$$

where U_i and each U_j are dependent on the actions of all n individuals. If $\theta = 1$ an individual gives equal weight to everyone in the group's utility as to his/her own utility, and behaves in whatever way is fully consistent with the cooperative solution to the prisoners' dilemma game.[5] If $\theta = 0$ the individual ignores the impact of his/her actions on others and behaves non-cooperatively. For those trained in economics this objective function may seem more plausible if we think of it as the objective function of a firm in an industry with n firms, and the U_i are profit functions. A $\theta = 1$ then corresponds to perfect collusion, $\theta = 0$ to Cournot independence, and $\theta = -1/(n-1)$ to the Bertrand equilibrium.

From the perspective of a rational egoist, only two values for θ can be justified. If the actions of the other individuals are contingent on i's actions, i.e., they cooperate if i does, then i sets $\theta = 1$. If not, then i sets $\theta = 0$, and ignores the consequences of his/her own actions on the utilities of others. If i is a pure rational egoist maximizing 0_i, it is very difficult to conceptualize a θ other than zero on one, a θ of 1/2 say.

But if we think of i as an individual responding, on the basis of prior experience, to the stimulus of being in a prisoners' dilemma situation, then θ could easily take on values between zero and one. What I propose is that we think of θ not as a parameter to be *chosen* by an individual, but as one which is characteristic of an individual or a group. People make decisions in prisoners' dilemma situations *as if* they were maximizing an objective function like (1). We can predict their behavior by deriving first and second order conditions from (1), and from these behavioral equations to be estimated from observable data. In prisoners' dilemma situations, our models will exhibit a better fit to the data if we allow θ to vary, than if we constrain it to zero, as is so often done in rational egoist models.

If we allow for the possibility of cooperative behavior in prisoners' dilemmas by assuming individuals maximize an objective function like (1), we can obviously proceed ahead and analyze (model) individual behavior with as much rigor and precision as we had without allowing for the possibility of

cooperation. Adding $\theta \sum_{j \neq i} U_j$ to the objective function complicates it, but does not alter its basic form. There is no sacrifice in rigor to choosing (1) as a maximand with $\theta > 0$.

One might grant all of this, but still resist bending from pure rational-egoism, on the grounds that explaining behavior with a model that allows $\theta > 0$ is *ad hoc*. Adding terms in θ to an equation may help increase the $\bar{R}^2$, but without a theory to explain θ, we really have not explained anything. To explain θ itself, however, we must step outside of the rational egoism paradigm. Let me explain why I believe a step into the realm of behaviorist psychology is not such a great leap as it might first appear.

4. Rational egoism and behaviorist psychology

The rational egoism postulate has two components: an assumption that individuals are egoistic, i.e., pursue their own ends, and an assumption that they do so in a rational or consistent manner. The first of these is certainly consistent with behaviorist psychology. Animals increase the frequency of those actions which are rewarded, reduce the frequency of those which are punished. The list of things which can serve as primary reinforcers is short and conforms to our notions of basic pleasures, e.g., food, drink, sex, as does the list of primary punishments (shock, extreme noise, heat, cold, light). If one can attribute motivation to observed behavior, then all animals appear to be pleasure-seeking, pain-avoiding creatures, hedonists of the most base kind. But so too do humans appear in our models of rational politics. With respect to postulated goals and aspirations, *homo economicus*, as usually seen in economic and public choice models, bears a close resemblance to Skinner's rat.

It is more difficult to reconcile the concept of rationality as it appears in rational egoist models, and the behavior of humans and animals described by psychologists. For example, a litmus test for rationality is whether an individual regards sunk costs as sunk, i.e., considers only the future benefits and costs of alternative actions, and ignores past (sunk) costs. Now, consider the results of the following experiments.

Experiment 1. Imagine that you have decided to see a play where admission is \$10 per ticket. As you enter the theater, you discover that you have lost a \$10 bill.

Would you still pay \$10 for a ticket for the play?

Experiment 2. Imagine that you have decided to see a play and paid the admission price of \$10 per ticket. As you enter the theater, you discover that you have lost the ticket. The seat was not marked and the ticket cannot be recovered.

Would you pay \$10 for another ticket?[6]

A rational individual would answer both questions yes (assuming a roughly constant marginal utility of money over the $20). If the play was worth the first $10, it should be worth the second. Moreover, the answer to both questions certainly should be the same, since a lost $10 bill is no more nor less sunk than a lost $10 ticket. But only 46 percent of the 200 students in experiment 2 answered yes compared to 88 percent of the 183 students in experiment 1. It would appear that many of the students in the second experiment felt that they had already spent $10 for a ticket, and regarded $20 as too much to spend for a single ticket. For these students, bygones were not bygones and their willingness to pay for a ticket was dependent on the context in which the decision had to be made, the way the question was 'framed.'[7] While the responses to these questions are difficult to reconcile with the behavior typically assumed of individuals in rational egoist models, they correspond to what we observe in everyday life. Individuals often do not ignore 'sunk costs' when making decisions, be it a businessman 'throwing good money after bad' trying to recoup unprofitable past investments in a doomed product line, or a President persisting to wage a lost war in Southeast Asia.

More generally, an effort to model human behavior based on realistic assumptions about 'how men think' as opposed to 'how they would think if they were rational' would place more emphasis on the experience of the individual in relationship to the context in which the decision is posed, and less on the consequences of the decision.

To many these distinctions may seem largely semantic. After all, even with animals we often describe behavior in rational egoist terms. We often speak, for example, of a dog as being intelligent. But by this we merely mean that the dog has learned well the tricks taught it. Sometimes we might say that a given dog is intelligent because it learns tricks more quickly than other dogs. But the dog still *learns* its tricks, and knowledge of the learning experiences of a dog is essential for predicting which tricks it will perform, and for inducing their performance.

We could say that it is rational for the thirsty rat to depress the lever if we are going to give it water for doing so. Indeed, some psychologists have taken to modeling the conditioned behavior of laboratory animals using concepts like marginal cost, marginal utility and objective function maximization (see, e.g., Staddon, 1983). Thus, we could defend the use of the rational egoism postulate alongside of the considerable evidence indicating the importance of prior conditioning (experience) in determining behavior, on the grounds that we are interested in modeling the behavior of only educated adults, well-trained dogs who know all the tricks. Their behavior appears rational, and it does not matter how they learned to behave as they do.

While this position can be defended *in certain contexts*, in others,

like the prisoners' dilemma, it does not allow us to develop hypotheses with which we can explain observed behavior. If noncooperative behavior is rewarded in a particular situation, how do we explain (predict) the cooperative behavior of some? Why does everyone not become a shoplifter as soon as the checkout clerk's back is turned?

The behaviorist psychology answer to this question is that not stealing is conditioned behavior, which we persist in even though in some situations it is not rewarded and noncooperative behavior would be. Moreover, certain stimuli induce us to recognize a prisoners' dilemma situation and thus elicit cooperative behavior, other stimuli do not. As consumers or producers acting in the market, we do not perceive ourselves to be in a prisoners' dilemma and thus generally behave noncooperatively.

The fact that different situations constitute different stimuli and thereby elicit different behavior helps explain some results in the experimental literature which are otherwise inexplicable. As noted above, the cooperation-noncooperation choice in a prisoners' dilemma is essentially the same as the collude-do not collude choice in an oligopolistic market. Both decisions can be modeled with equation (1). Given this correspondence, one would expect rational egoists to behave quite similarly in oligopoly-market experiments and voluntary contribution-public good experiments. But they do not. The perfect collusion-cooperative equilibrium in market experiments is a fragile flower that collapses into the noncooperative equilibrium at a whisper (Plott, 1982). Voluntary contributions at levels in excess of those implied by Nash-Cournot behavior are as common in public goods' provision experiments as noncooperative equilibria are in market experiments.[8] Explanation? The set-up of market experiments with demand schedules, cost schedules and the like suggests the kind of market environment in which individuals are used to behaving noncooperatively. The kind of environment in which noncooperative behavior has been rewarded. Thus, individuals act noncooperatively, unless they can be taught by punishments and rewards during the game to cooperate. Such intra-game learning of cooperative behavior typically occurs only when but two players are involved, and appears sensitive to the characteristics of the market.[9]

Some public good-voluntary contribution experiments of Isaac, McCue, and Plott (1985) are particularly revealing in this regard. They observed contributions above the noncooperative contribution on the first couple of trials of the experiment, but found a rather quick convergence to the noncooperative-free rider levels as the game was repeated with the same subjects. Their experiments differed from some other public good-free rider experiments in that 'the subjects were not allowed to communicate with one another during the experiment . . . [and] had no knowledge about the nature of any payoff charts other than their own' (p. 57). The context of the Isaac,

McCue, and Plott experiments was quite different from say, that of Schneider and Pommerehne (1981) in which students in a classroom had every reason to believe that they were in a real-life prisoners' dilemma game of the type in which cooperative behavior is expected, and in the past has been rewarded. Indeed, by conducting the experiment in the classroom, Schneider and Pommerehne may already have 'biased' the results toward cooperative behavior since students are accustomed to cooperating in a classroom in the numerous prisoners' dilemma situations they routinely encounter there. The inferences I draw from these experiments is that the first reaction of individuals when placed in a prisoners' dilemma situation is to cooperate to some degree. But individuals can be conditioned not to cooperate by repeating the game, if each individual is more or less isolated from the others as in the Isaac, McCue and Plott public good experiments, and most oligopoly experiments, and the noncooperative strategy is rewarded more heavily than is cooperation. Regardless of how one chooses to interpret these results one thing is clear, they totally contradict the behavior in prisoners' dilemmas predicted by the rational egoism postulate. Consistency with this postulate requires noncooperation if the game is played only once, but allows for cooperation to emerge in repeated games. What one observes, however, is the most degree of cooperation in experiments where the game is played only once, and a convergence to noncooperative, Nash equilibria upon repetition. The implications of these experiments for predicting behavior in actual prisoners' dilemma situations would seem to be that the extent of cooperative behavior will depend on the context of the game, and the conditioning experiences of the individuals playing it.

Behaviorist psychology can also help to explain another characteristic of the behavior observed in experiments, for which rational egoism has no explanation. Suppose one conditions a thirsty rat to press a lever by rewarding it with water after each lever press. Then one withholds the reward. One observes that the rat continues to press the lever for some time after positive reinforcement is withheld but at a diminishing rate *and intensity*. Eventually the operant behavior is extinguished entirely. Now consider a typical voluntary contribution-public good experiment. An individual is given a table of contributions and payoffs in which it is clear that the net payoffs to each group member are maximized if each voluntarily contributes \$10. It is also clear that if the contributions of the other members are independent of one's own contribution, the optimal contribution is zero. In addition to explaining a contribution of zero, rational egoism might explain a contribution of \$10 on the grounds that some individuals believe they are engaged in a supergame, and are employing a tit-for-tat strategy. Thus, rational egoism could be consistent with values of θ, the degree of cooperation in (1), for individuals of zero or one. But rational egoism cannot explain a value of θ for a single individual of say, 1/2, a contribution of \$5 in the public good

experiment. What rational strategy explains a half cooperative response? Reductions in both the frequency and intensity of cooperation are expected from behaviorist psychology when noncooperation is more highly rewarded than cooperation. Thus, behaviorist psychology is consistent with values of θ falling between zero and one, which is generally what one observes in voluntary contribution-public goods experiments (Marwell and Ames, 1981).

Behaviorist psychology principles have been used successfully to improve educational techniques, and in other contexts like mental hospitals. The behavior induced by reward systems in factories is also consistent with behaviorist psychology principles. The evidence that human behavior can be conditioned using the same principles developed to condition the behavior of rats, pigeons and other animals in a laboratory context is quite impressive.[10] I think it can also make a useful contribution to improving the predictive power of our models of economic and political behavior. One's first and most important learning experiences are in the home, in school, in church. One might expect individuals from stable home and school backgrounds, and with more intensive positive reinforcement for cooperative behavior and punishment for uncooperative behavior. These variables or proxies therefore might then be expected to be correlated with cooperative behavior in prisoners' dilemma contexts. By relying on behaviorist psychology we can add to our list of variables to explain crime, gift giving, voting and other types of behavior which pure rational-egoist models have generally been able to explain only poorly.[11]

I realize that for some the words 'behaviorist psychology' are themselves stimuli, which elicit a violent and negative response. I hasten to add, therefore, that one need not regard *Walden Two* as the most wonderful Utopia of which man has yet conceived, to believe that this methodology can serve as a useful complement to rational egoism when deriving testable hypotheses in some areas. We in public choice should be interested in using it to help us explain individual behavior in today's world with today's institutions. The more philosophical issues as to the good or evil that might ensue were behaviorist psychology used by a single authority to mold individual behavior need not deter us from employing these principles for a more mundane purpose.

5. Other areas in which rational egoism has weak predictive power

Although the focus of this article is on the breakdown of rational egoism in explaining behavior in prisoners' dilemmas, there are several other situations in which rational egoism runs into significant difficulty in accounting for human behavior, and psychology would appear to offer some

advantage. We briefly discuss but two.[12]

5.1 *Decisions involving risk, where probabilities are infinitesimal and pay-offs are very large*

Both Arrow (1982) and Heiner (1983) have noted the seemingly irrational behavior of many individuals when faced by decisions with very small probabilities of very large payoffs, citing the refusal of most individuals to purchase flood insurance at rates below its actuarial value as reported by Howard Kunreuther, et al. (1978). While this behavior, if it were rational, would imply an extreme willingness to take risks, other behavior with infinitesimal probabilities and large payoffs suggests extreme risk aversion, for example, when couples with children fly on separate planes to avoid the possibility that both parents would die in an airplane accident, or the measures taken by some recently to avoid contact with those with AIDS.[13] While information costs might explain some of these anomolies, as Arrow notes with respect to flood insurance, 'the information seems so easy to acquire and the stakes so large that this hypothesis hardly seems tenable' (1982: 2).

By definition, people cannot have much personal experience in making decisions in situations with infinitesimal probabilities and huge (i.e., often cataclysmic) payoffs. Thus, their behavior has not been conditioned by past punishment and rewards to resemble that of an individual consciously maximizing an egoist's objective function, as it does with more routine repetitive decisions. Instead, it resembles what it is, the almost random behavior of an individual confronted by a novel situation. Not surprisingly, chance events in an individual's past, like knowing someone who has the insurance, explain the decision to purchase insurance, where the variables rational egoism predicts, price and income, do not.

5.2 *Preference reversals*

A second form of inconsistency in decision making under risk, which has received considerable attention lately,[14] involves preference reversals. In these, individuals are first asked to choose between (state a preference for) two gambles, one with a high probability of winning a small sum of money, the other with a low probability of winning a considerably larger sum of money. Many individuals state a preference for the high probability-small sum gamble, but place a higher value on the low probability-high sum gamble. Preference reversals such as this example have been observed in a wide variety of contexts. These 'reversals can be seen not as an isolated phenomenon, but as one of a broad class of findings that demonstrate violations of preference models due to the strong dependence of choice and

preference upon information processing considerations' (Slovic and Lichtenstein, 1983: 597). When confronted by a multidimensional choice, individuals appear to *frame* the decision in such a way as to compare one dimension at a time (Tversky and Kahneman, 1981), ignoring small differences along a single dimension (Tversky, 1969). 'Whichever frame is adopted is determined in part by the external formulation of the problem and in part by the stardards, habits, and personal predilections of the decision maker' (Slovic and Lichtenstein, 1983: 600). Once again, to predict human behavior one needs information on the past experience of the individuals, and the context in which the decision is posed. Knowledge of the possible outcomes does not suffice.

6. The rational altruist – An alternative approach

My suggestion is that we can add realism and descriptive power to our modeling of human behavior if we retain the egoistic portion of rational egoism, and drop, or better modify the rationality assumption, at least in the strong form in which this assumption is usually employed. Before developing this argument, let me consider an alternative suggestion that is often made, that we drop the egoism assumption, in prisoners' dilemma situations, and assume that in these contexts the individual adopts a set of ethical preferences. That is, that we assume that man has two natures: a selfish nature, and a cooperative-altruistic nature, and that he makes decisions in some contexts using his selfish nature, and in others using the altruistic-cooperative nature.[15] This Jekyll and Hyde view of man has been around at least since Plato. More revent variants on it would certainly include Goethe's *Faust*, Oscar Wilde's *Portrait of Dorian Gray*, as well as Robert Louis Stevenson's classic tale. Can all of these great intellects have been wrong? Only a fool would answer yes. One can, of course, employ a Jekyll and Hyde view of man to reconcile cooperative and noncooperative behavior with rationality. Indeed, one can use equation (1) to model it. Mr. Hyde sets θ equal to zero, Dr. Jekyll to one. Now for the moral philosopher these are the only two possible values which are admissable. One either cooperates or one does not. The ethical rule, 'Do for others half as much as you would have them do for you,' cannot be countenanced. But this Jekyll-Hyde view of man as a *description* of behavior is certainly rejected by the data. As an explanatory hypothesis, it would have to be weakened. But then how does one predict when an individual will be Hyde, when Jekyll? How does one explain the Jekyll-like behavior of an individual in some prisoners' dilemma situations, Hyde-like behavior in others? What explains partial cooperative behavior? If one attempts to convert the Jekyll-Hyde view of man into a predictive theory, one has to begin considering

factors such as those suggested by behaviorist psychology.

The advantage of starting with behaviorist psychology is that it allows us to begin with a unified view of human behavior. Man is basically base. Behaviorist psychology's premises regarding human behavior are fully compatible with the egoistic portion of rational egoism. Although base, man can be taught to behave nobly, to cooperate. Moreover, the principles by which this behavior is learned can be studied and used to predict cooperative behavior. Hypotheses can be derived and tested with experimental data using standard statistical tests. Going to behaviorist psychology is less of a methodological leap for a social scientist who works with rational egoist models than going to some competing sociological-psychological theories.

That cooperation can be explained as egoistic behavior is perhaps not so surprising, since individuals are rewarded for cooperating when all participants in a prisoners' dilemma cooperate. What presents a greater challenge for a behavioral theory based on pure egoism assumption is altruism, 'behavior carried out to benefit another without anticipation of rewards' (Macauley and Bercowitz, 1970: 73). Altruism would appear to offer the greatest promise for a theory of behavior based on the existence of a set of ethical (unselfish) preferences.

The first problem one confronts in analyzing altruism as unselfish behavior is the hedonistic paradox. 'Even the most unselfish act may produce a psychological reward for the actor' (Cohen, 1978: 82–83). Ronald Cohen (1978) observes that the forms altruism takes vary across cultures, and from this suggests 'that altruism is [psychologically] culture-dependent and not part of some theoretical notion of human nature' (p. 92). Nor does altruism in man appear to be an inherited behavioral trait. If Darwinian forces play a role in inducing altruistic behavior, then they would appear to do so by shaping the cultural traditions which foster certain forms of altruism. 'Sociocultural phenomena persist through time because they have survival value; that is, they tend to aid in group survival more adequately than does some other set of contending traditions. The presence and persistence of altruistic values or elements of altruism have survival value for the group that maintains such beliefs . . . there is in human nature no genetic basis for altruism . . . except for some biological capacity to learn it. Whether or not altruism exists, and to what extent, lie in the nature and evolution of the sociocultural system, which then in turn has effects on the motivations and behaviors of individuals' (Cohen, 1978: 93).[16]

Although these arguments do not preclude one from postulating a set of ethical preferences to explain altruistic behavior, the application of Occam's razor dictates maintaining a purely egoistic assumption regarding human behavior if that suffices, as would appear to be the case. As Cohen puts it, 'the evidence on man as a hedonistic, self-gratifying creature is so strong, that I would rather interpret giving as a social act and a cultural tradition'

(1978: 94). Thus, 'the hedonistic paradox makes a good logical basis for postulating man's essentially hedonistic nature' (Cohen, 1978: 83).

7. Social institutions and cooperative behavior

Although the prisoners' dilemma supergame is not a useful analytic tool for *describing* individual behavior in many actual prisoners' dilemma situations, it is useful *normative* construct for guiding society. At the societal level, we are all engaged in an indefinite unbounded series of prisoners' dilemma games with respect to stealing and the like. While it is sometimes in an individual's interest not to cooperate in a prisoners' dilemma, it is at the same time in the interest of society to punish noncooperation and reward cooperation in these situations. Taboos, social sanctions, mores and more formal institutions to induce cooperative behavior evolve, as the previous discussion of altruism suggests, to further the collective interests of the group.

At the group level, then, egoism leads us to establish institutions which condition individuals to maximize an objective function like (1) in prisoners' dilemma situations with θ equal to one. Thus, in prisoners' dilemma situations (1) can be viewed as both a description of individual behavior with θ free to vary between zero and one, a parameter to be determined by the data; and (1) can be viewed as a normative objective function for the group in designing its educational and other cooperation conditioning institutions. Note that to the extent we are able to build institutions which induce individuals to maximize (1) with $\theta = 1$ in prisoners' dilemma situations, we induce each individual to behave in these situations so as to maximize a Benthamite social welfare function defined over the group.

8. Cooperative behavior and sociobiology

My message is in part that we should expand our frame of reference in some circumstances beyond the individual maximizing his/her utility ignoring the consequences of this behavior on others. Psychology gives us an explanation for why other factors may explain individual behavior in certain contexts and what these other factors might be. But, although encompassing individual behavior to some extent, principles of psychology must also be viewed as but a subset of a broader set of laws governing behavior.

A pigeon, for example, can be conditioned to peck a key for food, but it cannot be trained to peck a key to avoid receiving a shock. For a pigeon, pecking and eating are sufficiently closely related, so that food can be used as a reinforcer to condition pecking. Shocks cannot. What behavior is

subject to operant conditioning and through what rewards and punishments appears to be genetically determined. Our environment and genetic heritage constrain the kinds of behavior which we can learn, and the speed of learning (Staddon, 1983: 12). The laws of evolution in turn determine which members of a species, or groups of members survive, which genes pass on. Evolution selects those gene structures that maximize survival chances. Here then we have yet another equation to be maximized to explain individual behavior.[17] In terms of equation (1), we must recognize that the U's themselves are endogenous variables with respect to the biological processes which determine man's evolution. Social animals may evolve gene structures which facilitate learning cooperative behavior, and which make teaching cooperative behavior to one's offspring a quasi-instinctual act.[18] But cooperative behavior still must be taught, and this teaching-learning process follows definite laws. Sociobiology defines the boundaries in which learning takes place, the modes of cooperative behavior which can and will be taught, but it does not guarantee the teaching.

10. Summary and conclusions

What I have proposed for consideration is a view of human behavior which is a complement to both the rational egoist model of individual behavior, and the sociobiological model of species behavior. Cooperative behavior can benefit all members of a group and increase each individual's chances of survival. Rather complicated cooperative behavior patterns are observed in some species, and give testimony to the importance of genetic heritage in explaining behavior. But anyone who has ever observed small children at play must believe that the instincts for selfish action dominate those to cooperate in humans.[19] What humans inherit is a capacity to learn, to discriminate one situation from another, to generalize from past experience. But cooperative behavior among humans is learned.

The usual depiction of this learning process by modelers of rational egoistic behavior is to assume that learning takes place within the context of the game. Since learning takes time, a repeated game is required to achieve cooperation. The most successful strategy for teaching one's fellow-player to cooperate appears to be the tit-for-tat strategy. It is interesting how closely Axelrod's (1984) description of tit-for-tat's success in bringing about cooperation parallels the psychologist's description of learning. The cooperative behavior of one player is rewarded by the cooperative behavior of the other. Noncooperative behavior is punished. 'The emergence, growth, and maintenance of cooperation ... require an individual to be able to recognize another player who has been dealt with before [read stimulus]. They also require that one's prior history of interactions with

this player can be remembered, so that a player can be responsive' [read one has a conditioned behavior pattern] (p. 174).

Axelrod emphasizes that 'there is no need to assume that the players are rational. They need not be trying to maximize their rewards. Their strategies may simply reflect standard operating procedures, rules of thumb ... habits ... The actions that players take are not necessarily even conscious choices' (p. 18, and again at p. 173). Axelrod thus clearly believes that the tit-for-tat strategy as a description of behavior applies to situations, in which the rational portion of the rational egoism assumption is inappropriate. Throughout the book he moves back and forth from examples for which *rational* egoism in a prisoners' dilemma might fit, and others where it will not. He spends a whole chapter on 'The Evolution of Cooperation in Biological Systems' (written with William D. Hamilton, Chapter 5). The parallel between the basic principle of the evolutionary approach, 'whatever is successful is likely to appear more often in the future' (p. 169), and the basic principle of behaviorist psychology, 'positive reinforcement increases the future likelihood of operant behavior,' is again obvious. Operant conditioning describes the learning process of the *individual* in adapting to the environment, natural selection describes the evolutionary process of *species* adaptation (Notterman, 1970: 13).

The common thread running through each of these paradigms from selfish gene, through thirsty rat, to *homo economicus* is egoism. And, I submit, the only assumption essential to a descriptive and predictive science of human behavior is egoism. What then of the rationality assumption? A moment's reflection will reveal that the only use to which we put the rationality assumption is to add precision to the predictions stemming from the egoism postulate, as when we operationalize rational egoism by assuming the individual maximizes an objective function. Now rigor is an important and useful property of a model, so long as it does not come at the expense of realism. Fortunately, the public choice analyst can have both the rigor of mathematical modeling and the realism of assuming only egoistic motivation, if he/she treats the rationality assumption as an *as if* behavioral assumption.

Over 35 years ago, Armen Alchian (1950), answered critics of the profits maximization assumption in the theory of the firm by arguing that in a competitive environment the less profitable firms perish, and the surviving firms adopt decision rules *as if* they had been consciously trying to maximize profits, whatever the criteria actually employed to make decisions. Competition for survival selects gene structures *as if* the evolution of the species were maximizing the probability of survival. Where cooperation has significant advantages for all members of a group, the group will adopt mores and laws which condition people to behave *as if* they were maximizing a function like (1) with $\theta = 1$. Over time, those social institutions for conditioning

cooperative behavior will survive that maximize group survival chances. The weaker the social conditioning of cooperative behavior is, the more individual behavior will resemble the maximization of (1) with some $\theta < 1$. In the limit, when individual decisions do not impinge of the welfare of others, individual behavior resembles the maximization of (1) *as if* $\theta = 0$. Thus, behavior can be modeled *as if* it was motivated by conscious choices to maximize different objective functions, with the specification of the objective function varying to reflect the level of analysis (species, group, individual), the adaptive history of the subjects, and the context of the actions.

For most of us this is a comforting thought, since we have been conditioned to working with optimization models. But there is an important difference between *as if* maximization based on a pure egoism postulate and rational egoism as usually employed. The difference lies in determining what goes into the objective function.

A model of man based on *as if* maximization of a socially conditioned or an evolutionarily molded objective function, is a model of *adaptive* behavior. Thus, what I am proposing is that the rational egoism postulate be replaced by an adaptive egoism postulate to model human behavior. The difference between these assumptions is similar to the difference between the rational expectations and adaptive expectations assumptions used in economics. To describe an individual's expectations at time t using the rational expectations assumption we look at what happens after t, to describe them under adaptive expectations we look at what happened before t. Rational egoism, as usually modeled, considers only the consequences of different actions, the payoffs in the strategy matrix. When modeling the behavior of an individual, who has played the game several times before, who has been rewarded for choosing certain strategies, and punished for others, this forward looking objective function can give accurate predictions of an individual's actions. For much of what we do in public choice as in economics, this rarefied setting for making choices may come close enough to matching the context in which isolated individuals act to give us reasonable descriptions of human behavior. But when we attempt to explain behavior in more complicated social contexts, as we so often do in public choice, charity, voting, crime, then we shall add to the descriptive power of our models by recognizing that man has not only a future, but also a past.

George C. Homans (1958, 1962, 1964, 1967, 1974) has argued for many years that a unified sociological theory can and should be built on behavioralist psychology principles to displace the many discordant theories of modern sociology.[20] Were economics and rational politics to replace rational egoism with adaptive egoism as their fundamental behavioral postulate, the possibility would exist for constructing a common methodological foundation for all of the social sciences. For once

interdisciplinary research within the social sciences would not be hampered by basic differences in methodologies. The potential gains in knowledge from cross-fertilization are enormous. Certainly these gains are worth the costs of revising our modeling of human behavior, where necessary, as the replacement of rational egoism by adaptive egoism dictates.

NOTES

1. The most recent defense of the prisoners' dilemma supergame explanation for cooperation is by Robert Axelrod (1984). But Axelrod's entire book is based on the analysis of 'interactions between just two players at a time' (p. 11). For further discussion of the incentive to defect in an *n*-person prisoners' dilemma, see Taylor (1976): 43–61), and Hardin (1982: 42–49).
2. (1955: 450) Quotation taken from Williamson (1985: 2–3)
3. See also Witt (1986) for a critique of the rational egoism modeling for behavior in prisoners' dilemma situations.
4. See standard books on psychology like Notterman (1970), Schwartz and Lacey (1982), and Staddon (1983).
5. Cooperation can be modeled in other ways, as for example in Margolis (1982), but assuming a single objective function with different weights on own and other utilities seems to me to be the simplest way. For a still different approach, see Kliemt (1984).
6. Taken from Tversky and Kahneman (1981: 457).
7. Ibid.
8. See Chamberlin (1978), Marwell and Ames (1979, 1980, 1981), Smith (1979a, b), and Schneider and Pommerehne (1981).
9. Plott (1982), Hoffman and Spitzer (1985: 1003–1005). Phillips, Battalio, and Holcomb, Jr. (1985) find that the ability of duopolists to cooperate is strongly influenced by the history of the market. Alger (1986) finds more cooperative behavior than Plott's (1982) survey implies and, surprisingly, that the number of sellers does not prove to be an important variable in determining performance (over the range of 2–4 sellers).
10. See, e.g., Millenson (1967), Notterman (1970), Schwartz and Lacey (1982: Chs. 2–6, and Staddon (1983).
11. Similar sentiments have been expressed by Witt (1985, 1986).
12. A third, relying on the psychologists' concept of cognitive dissonance, is examined by Akerlof and Dickens (1982).
13. Ferejohn and Fiorina (1974) attempt to rationalize the seemingly irrational act of voting in the face of an infinitesimal probability that a single vote is decisive by arguing that voters choose a minimax regret strategy. As Mayer and Good (1975) among others point out, such a choice implies an implausible (irrational) degree of risk aversion on the point of voters. Moreover, assuming a high overlap between voters and nonbuyers of flood insurance, one or the other of these decisions must be irrational or individuals veer wildly from extreme risk aversion to extreme risk taking as they move from one risky choice to another. Such behavior, even if labeled rational, forms no basis for a predictive model of human behavior.
14. See Grether and Plott (1979), Pommerehne, Schneider, and Zweifel (1982), and Slovic and Lichtenstein (1983).
15. This suggestion for improving the predictive (positive) performance of our models has been put forward most recently by Margolis (1982) and Etzioni (1986). A similar suggestion in a normative context was made by Harsanyi (1955); see also the discussion by Arrow

(1963: 81–91).

16. Egalitarian sharing rules would also appear to be prime candidates for explication by a rational equity theory. Yet they too can be explained by assuming that individuals are selfish and equity rules maximize group rewards (Hatfield, Walster, and Piliavin, 1978).
17. Jack Hirschleifer (1977, 1980, 1985) has emphasized the parallel between the maximization assumption as used in economics, and the model of behavior underlying sociobiology.
18. See Hirschleifer (1980).

    Whether evolution favors genes which promote cooperative group behavior or selfish individual behavior is a complicated and unsettled question (Wilson, 1975; Becker, 1976; Hirschleifer, 1977; Margolis, 1982: 26–27, 31–35). While the consensus appears to be that when noncooperative behavior favors individual survival probabilities over group survival, it is the noncooperative genes which will survive. But the arguments and evidence for this position seem to be based too heavily on nonhuman species and thus on inherited cooperative or noncooperative behavior.

    The salient feature of man's genetic evolution is the brain's development, and the capacity it provides to learn and remember. The development of culture brings with it mores and social sanctions, which punish behavior that otherwise would benefit the individual at the group's expense. Sanctions lowering the survival chances of a noncooperative individual's genes (banishment, incateration, dismemberment, execution) have been used by social groups for as far back as the word 'social' has any descriptive content.
19. The instinctual tendency to emit cooperative may be stronger at subsequent stages of development, however. See Kohlberg (1963) and Staddon (1983: 14).
20. See also Viktor Vanberg (1975, 1983) and Hans J. Hummell and Karl-Dieter Opp (1971).

REFERENCES

Alchian, A.A. (1950). Uncertainty, evolution and economic theory. *Journal of Political Economy* 58 (June) 211–221.

Alger, D. (1986). *Investigating oligopolies within the laboratory*. Washington: FTC.

Arrow, K.J. (1963). *Social choice and individual values*, rev. ed. New York: Wiley.

Arrow, K.J. (1982). Risk perception in psychology and economics. *Economic Inquiry*: 20 (January) 1–9.

Axelrod, R. (1984). *The evolution of cooperation*. New York: Basic Books.

Becker, G.S. (1976). Altruism, egoism, and genetic fitness. *Journal of Economic Literature*: 14 (September) 817–826.

Bridgeman, P. (1955). *Reflections of a physicist*, 2nd ed. New York: Philosophical Library.

Chamberlin, J.R. (1978). The logic of collective action: Some experimental results. *Behavioral Science* 23: 441–445.

Cohen, R. (1978). Altruism: Human, cultural, or what? In L. Wispe (Ed.), *Altruism, sympathy, and helping*, 79–98. New York: Academic Press.

Coughlin, P. (1982). Pareto optimality of policy proposals with probabilistic voting. *Public Choice* 39(3): 427–433.

Coughlin, P. and Nitzan, S. (1981). Electoral outcomes with probabilistic voting and Nash social welfare maxima. *Journal of Public Economics* 15 (February): 113–121.

Etzioni, A. (1986). *The case for a multiple utility conception*. George Washington University. Mimeo.

Ferejohn, J.A., and Fiorina, M.P. (1974). The paradox of not voting: A decision theoretic analysis. *American Political Science Review* 68: (June) 525–536.

Grether, D.M., and Plott, C.R. (1979). Economic theory of choice and the preference reversal phenomenon. *American Economic Review* 69 (September): 623–638.

Hardin, R. (1982). *Collective action*. Baltimore: Johns Hopkins University Press.

Harsanyi, J.C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy* 63 (August): 309–321.

Hatfield, E., Walster, G.W., and Piliavin, J.A. (1978). Equity theory and helping relationships. In L. Wispe (ed.), *Altruism, sympathy, and helping*. New York: Academic Press. 115–139.

Heiner, R.A. (1983). The origin of predictive behavior. *American Economic Review* 73 (September): 560–595.

Hirschleifer, J. (1977). Economics from a biological viewpoint. *Journal of Law and Economics* 20 (April): 1–52.

Hirschleifer, J. (1980). Privacy: Its origin, function, and future. *Journal of Legal Studies* 9 (December): 649–665.

Hirschleifer, J. (1985). The expanding domain of economics. *American Economic Review* 75 (December): 53–68.

Hoffman, E., and Spitzer, M.L. (1985). Experimental law and economics: An introduction. *Columbia Law Review* 85 (June): 991–1036.

Homans, G.C. (1958). Social behavior as exchange. *American Journal of Sociology* 62 (May) 597–606.

Homans, G.C. (1962). *Sentiments and activities: Essays in social science*. New York: Free Press.

Homans, G.C. (1964). Bringing men back in. *American Sociological Review* 29 (December) 809–818.

Homans, G.C. (1967). *The nature of social science*. New York: Harcourt, Brace & World.

Homans, G.C. (1974). *Social behavior: Its elementary forms*, rev. ed. New York: Harcourt, Brace & World.

Hummell, H.J., and Opp, K.-D. (1971). *Die Reduzierbarkeit von Soziologie auf Psychologie*. Braunschweig: Vieweg.

Isaac, R.M., McCue, K.F., and Plott, C.R. (1985). Public goods provision in an experimental environment. *Journal of Public Economics* 26 (February): 51–74.

Kliemt, H. (1984). Nicht-explanative Funktionen eines Homo Oeconomicus' und Beschränkungen seiner explanativen Rolle. In M.J. Holler (ed.) *Homo Oeconomicus II*, 7–49. Munich: N. Leudemann.

Kohlberg, L. (1963). 'The development of children's orientations toward a moral order, I. Sequence in the development of moral thought. *Vita Humana*, 6: 11–33.

Kunreuther, H., et al. (1978). *Disaster insurance protection*. New York: Wiley.

Ledyard, J.O. (1984). The pure theory of large two candidate elections. *Public Choice* 44(1): 7–41.

Macauley, J.R., and Berkowitz, L. (1970). *Altruism and helping behavior: Social psychological studies of some antecedents and consequences*. New York: Academic Press.

Margolis, H. (1982). *Selfishness, altruism, and rationality*. Cambridge: Cambridge University Press.

Marwell, G., and Ames, R.E. (1979). Experiments on the provision of public goods I: Resources, interest, group size, and the free rider problem. *American Journal of Sociology* 84 (May): 1335–1360.

Marwell, G., and Ames, R.E. (1980) Experiments on the provision of public goods II: Provision points, stakes, experience and the free rider problem. *American Journal of Sociology* 85: 926–937.

Marwell, G., and Ames, R.E. (1981) Economists free ride, does anyone else? *Journal of Public Economics* 15 (June): 295–310.

Mayer, L.S., and Good, I.J. (1975). Is minimax regret applicable to voting decisions? *American Political Science Review*. 69 (September): 916–917.

Millenson, J.R. (1967). *Principles of behavioral analysis*. New York: Macmillan.

Notterman, J.M. (1970). *Behavior: A systematic approach*. New York: Random House.

Phillips, O.R., Battalio, R.C., and Holcomb, J.H., Jr. (1985). *Observations on history and market signalling in duopoly markets*. University of Wyoming. Mimeo.

Plott, C.R. (1982). Industrial organization theory and experimental economics. *Journal of Economic Literature* 20 (December): 1485–1527.

Pommerehne, W.W., Schneider, F., and Zweifel, P. (1982). Economic theory of choice and the preference reversal phenomenon: A reexamination. *American Economic Review* 72 (June): 569–574.

Schneider, F., and Pommerehne, W.W. (1981). On the rationality of free riding: An experiment. *Quarterly Journal of Economics* 96 (November): 689–704.

Schwartz, B., and Lacey, H. (1982). *Behaviorism, science, and human nature*. New York: Norton.

Slovic, P., and Lichtenstein, S. (1983). Preference reversals: A Broader perspective. *American Economic Review* 73 (September): 596–605.

Smith, V.L. (1979a). Incentive compatible experimental processes for the provision of public goods. V.L. Smith, (ed.), *Experimental economics*, vol. 1, 59–168. Greenwich: JAI Press.'

Smith, V.L. (1979b). An experimental comparison of three public good decision mechanisms. *Scandinavian Journal of Economics* 81(2): 189–215.

Staddon, J.E.R. (1983). *Adaptive behavior and learning*. Cambridge: Cambridge University Press.

Taylor, M. (1976). *Anarchy and cooperation*. New York: Wiley.

Tullock, G. (1981).; Why so much stability? *Public Choice* 37(2): 189–202.

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review* 76 (January): 31–48.

Tversky, A., and Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science* 211 30 (January): 453–458.

Vanberg, V. (1975). *Die zwei Soziologien*. Tübingen: J.C.B. Mohr (Paul Siebeck).

Vanberg, V. (1983). 'The rebirth of utilitarian sociology. *Social Science Journal* 20 (July): 71–78.

Williamson, O.E. (1985). *The economic institutions of capitalism*. New York: Free Press.

Wilson, E.O. (1975). *Sociobiology: The new synthesis*. Cambridge: Belknap.

Witt, U. (1985). Economic behavior and biological evolution: Some remarks on the sociobiology debate. *Zeitschrift für die gesamte Staatswissenschaft* 141 (September): 365–389.

Witt, U. (1986). Evolution and stability of cooperation without enforcable contracts. *Kyklos* 38(2).