ELSEVIER

Ecological Economics 35 (2000) 311–322

ECOLOGICAL ECONOMICS

www.elsevier.com/locate/ecolecon

SPECIAL ISSUE

THE HUMAN ACTOR IN ECOLOGICAL-ECONOMIC MODELS

Beyond *Homo economicus*: evidence from experimental economics

Herbert Gintis [1]

Department of Economics, University of Massachusetts, Amherst, MA 01003, USA

Abstract

Environmental policies are generally based on a model of the human actor taken from neoclassical economic theory. This paper reports on laboratory experiments suggesting weaknesses in this model and describes alternative models correcting these weaknesses. One finding is that economic actors tend to be *hyperbolic* as opposed to *exponential* discounters who discount the immediate future at a higher rate than the more distant future. Another finding is that economic actors are not self-regarding, but rather in many circumstances are *strong reciprocators* who come to strategic interactions with a propensity to cooperate, respond to cooperative behavior by maintaining or increasing cooperation, and respond to free-riders by retaliating against the 'offenders', even at a personal cost, and even when there is no reasonable expectation that future personal gains will flow from such retaliation. We discuss some implications for policy analysis. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Hyperbolic discounting; Economic theory; Expected utility principle

1. Introduction

Balancing the goal of improving the natural environment against other *desirata*, such as increased consumption and leisure, is a problem of 'marshalling scarce resources towards competing ends', to use the well-known phrase of Lionel Robbins (1935). Economists have amassed an impressive and sophisticated body of economic theory to deal with such issues. However, this theory has been developed and tested in the context of goods and services that are at least in principle capable of being priced, and over which individuals have operationally meaningful preferences. Such concerns as depletion of the ozone layer, reduction in biodiversity, and the destruction of rain forests, to mention only a few of the more prominent environmental concerns, are sufficiently removed from the sorts of issues with

[1] The author is Professor of Economics, University of Massachusetts, Amherst. This paper is loosely based on Chapter 11 of Herbert Gintis, Game Theory Evolving (Princeton University Press, 2000). The author can be contacted at *gintis@econs.umass.edu*and related material can be found at the web site *http://www-unix.oit.umass.edu/ ~ gintis*

0921-8009/00/$ - see front matter
PII: S0921-8009(00)00216-0

which economists have traditionally dealt that it would be surprising if this body of theory did not require serious revisions to deal with environmental policy. And it does, indeed, require at least one such emendation, in the form of an expanded and revised model of the individual economic actor.

The standard model of the individual in economic theory, whom we shall call *Homo economicus*, has several characteristics that are relatively unproblematic in a market setting, but have potentially seriously misleading implications when applied outside this sphere. *H. economicus* comes to a choice situation with exogenously given and determinate preferences. These preferences apply to goods and services that are produced, consumed, and exchanged. *H. economicus* is self-interested, caring only about personal (or more broadly, familial) bundle of commodities, work, and leisure acquired. *H. economicus* is outcome-oriented, caring about social interactions only insofar as they affect his final consumption and wealth. Finally, *H. economicus* has a rate of time preference that allows him to allocate consumption over time in a consistent manner, reflecting his welfare and his concern for the welfare of future generations.[2]

Everyday observation attests to the fact that people sometimes fail to conform to this model. People succumb to harmful temptations, behave charitably and/or vengefully, and have a concern for fairness. Economists have placed little credence in such observations on the grounds that in a natural setting behaviors appearing to conform to such descriptions could simply be rational self-interested behavior in the presence of reputation effects, asymmetric information, and unusual preferences. Some examples may make this point clear.

1. Drug addiction may seem a perfect example of people making choices that are not in their self-interest. However, a much larger fraction of those who try drugs either give them up or maintain their use at recreational levels than become addicted. Therefore drug taking may be a risky behavior the net benefit of which is positive, even though it has a negative payoff for some.
2. Most charitable giving is not anonymous, and self-interested agents may act charitably to cultivate a reputation that benefits them with their family, neighbors, and coworkers. Similarly, acts of selfless heroism may give signals to others that ultimately materially benefit the hero.
3. Acts of vengeance are common, but such acts also serve the function of establishing the vengeful individual as someone who must be treated with great care. The cases where vengeance is taken to an excess that cannot plausibly be in the perpetrator's interest may simply be judgmental errors, or a mark of mental disorder afflicting a small fraction of people.
4. Generous acts towards neighbors, coworkers, and friends may be forms of self-interested reciprocity — called 'tit-for-tat' in the game-theoretic literature (Axelrod, 1984) and 'reciprocal altruism' (a misnomer, since it is not true altruism) in the biological literature (Trivers, 1971).
5. What seems like excessively risky behavior, or excessively present-oriented behavior, may simply reflect the individuals personal predilection for weighing future against present costs and benefits. Since there is no obvious objective means of determining the 'correct' rate of time preference, this behavior remains within the purview of *H. economicus*.

However, the development of axiomatic decision theory and more recently game theory have provided a rigorous context for distinguishing among alternative hypotheses concerning human behavior, by moving from everyday life to laboratory and field experiments allowing for a relatively controlled environment for testing alternative models of behavior. These theoretical constructs provide the tools for carefully modeling the conditions of social interaction, the characteristics of players, the rules of the game, the informational structure available to the participants, and the payoffs associated with particular decisions and

[2] Neoclassical economics prefers the term 'rational actor' to *H. economicus*. We avoid this term because there is nothing particularly 'rational' (or for that matter 'irrational') about his behavior.

strategic interactions. They have thus fostered the growth of *experimental economics* as a means of modeling human behavior.

While many of the predictions of traditional economic theory have been verified experimentally, many others have been decisively disconfirmed. I shall present here a body of evidence flowing from such experiments suggesting that the *H. economicus* model is incorrect outside of an anonymous market setting. I shall also make reference to a growing body of analytical models of human choice that are based on a broader and more accurate set of assumptions. First, we find that human decision-making systematically violates the axioms of decision theory. In particular, individuals betray a *status-quo bias*, and are *hyperbolic* rather than *exponential* discounters of benefits and costs that accrue in the future. Second, where agents can engage in strategic interaction, with the power to reward and punish the behavior of other players, the predictions of game theory generally fail. Second, actors in economic settings involving strategic interaction are cooperative and prosocial in a way not predicted by *H. economicus*. In particular, economic actors in many circumstances behave more like *Homo reciprocans* than *H. economicus*: they are *strong reciprocators* who come to strategic interactions with a propensity to cooperate, respond to cooperative behavior by maintaining or increasing cooperation, and respond to noncooperative free-riders by retaliating against the 'offenders', even at a personal cost, and even when there is no reasonable expectation that future personal gains will flow from such retaliation.

2. Experiments in individual choice behavior

A 'game against nature' involves a single agent choosing under conditions of uncertainty. For instance, uncertainty may be due to a random draw or a natural event (crop loss, death). A 'game against oneself' is a choice situation in which an agent optimizes over time, but cannot automatically precommit to carrying out in the future the plans being made in the present. In this section we present the results of laboratory studies of games against nature and ourselves.

2.1. Time inconsistency and hyperbolic discounting

'Time consistency' means that the future actions required to maximize the current present value of utility remain optimal in the periods when the actions are to be taken. The central economic theorem on choice over time is that time consistency requires that the future be discounted at a fixed rate, independent of when the costs and benefits of the actions actually occur. People tend not be time consistent. Rather, they appear to have higher discount rates over payoffs in the near future than in the distant future. It follows that people often favor short-term gains that entail long-term losses. We often term this 'impulsivity' or 'weakness of will'. It follows that traditional benefit-cost analysis may underestimate the long-term benefits of environmental policies by discounting distant payoffs at too high a rate.[3] Moreover, people do not have the same discount rate for all types of payoffs. In particular, in some circumstances laboratory subjects consistently exhibit *negative* discount rates.

Take, for instance, impulsive behavior. Economists want to argue that what appears to be 'impulsive' — such as cigarette smoking, drug use, unsafe sex, overeating — may in fact be welfare-maximizing for people who have high time discount rates or who prefer acts with high future costs. Controlled experiments in the laboratory cast doubt on this explanation, indicating that people exhibit a *systematic tendency to discount the near future at a higher rate than the distant future* (Chung and Herrnstein, 1967; Loewenstein and Prelec, 1992; Herrnstein and Prelec, 1992; Fehr and Zych, 1994; Kirby and

[3] It has been argued that the very concept of discounting future benefits and costs is misguided. This position implies that benefits and costs should not be taken into account in assessing environmental policy, since without a discount rate (even zero, or negative), there is no way to aggregate costs and benefits across time. The fact that economic agents do not have time-consistent discount rates by no means justifies so extreme a position. Rather, it suggests that we cannot infer the terms according to which people are willing to make long-term tradeoffs from the risk-free interest rate or other measure of short-term discount rates.

Herrnstein, 1995). In fact, observed intertemporal choice appears to fit the model of *hyperbolic discounting* (Ainslie and Haslam, 1992; Ainslie, 1975; Laibson, 1997), first observed by Richard Herrnstein in studying animal behavior (Laibson and Rachlin, 1997).

In addition, agents have different rates of discount for different types of outcomes (Loewenstein, 1987; Loewenstein and Sicherman, 1991). For instance, suppose a subject in an experiment is offered two free dinners for oneself and a friend, one at a local fast-food restaurant, and one at the finest restaurant in the region. The subject is told he or she may take either one next weekend, and the second in 2 months from now. Economic theory tells us that with positive discount rates, the subject should take the expensive dinner now and the inexpensive one later. But in fact, most subjects opt for the inexpensive dinner before the expensive! When asked why, the subjects report that they enjoy the *anticipation* of having the really good meal (Loewenstein, 1987). In a similar experiment (Loewenstein and Sicherman, 1991), it is shown that, all else being equal (e.g. the total sum of payoffs is held constant), workers prefer increasing over declining wage profiles. This again can happen only if workers have a negative time discount rate. When asked why they chose as they did, subjects said that they considered pay increases as a reward for hard work, and that job motivation would be difficult to maintain with declining wages.

The implication for economic policy is that care must be taken to elicit valid long-term discount rates of voters and citizens for the particular policies and explicit social situations under consideration, rather than following the standard practice of discounting at the rate given by short term risk-free interest rates on financial markets, which are likely to be excessively high. This point is especially important for environmental policy since the benefits of conservation often lie far in the future and hence are treated as being virtually worthless in traditional benefit-cost analysis, which uses a discount rate in the range of 3% per year. Such a discount rate reduces a benefit that accrues in 100 years to about 4.75% of its future value, whereas at a discount rate of 1%, the present value is 36.6, which is about 7.1 times as great. At a zero discount rate, the present value is, of course the same today as in 100 years, which is thus about 21 times as great as with a 3% discount rate.

2.2. Choice under uncertainty: logic and heuristics

The centerpiece of the theory of choice under uncertainty is the *expected utility principle*, which says that 'rational' agents choose among uncertain payoffs to maximize the expected utility of the payoffs. Von Neumann and Morgenstern (1944), Friedman and Savage (1948), Savage (1954) and Anscombe and Aumann (1963) showed that the expected utility principle can be derived from the assumption that agents have consistent preferences over an appropriate set of uncertain payoffs.

Laboratory testing of the *H. economicus* model of choice under uncertainty was initiated by the psychologists Daniel Kahneman and Amos Tversky. In a famous article in the journal *Science*, Tversky and Kahneman (1974) summarized their early research as follows:

> How do people assess the probability of an uncertain event or the value of an uncertain quantity?...people rely on a limited number of heuristic principles which reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations. In general, these heuristics are quite useful, but sometimes they lead to severe and systematic errors.

Subsequent research has strongly supported this assessment (Kahneman et al., 1982; Shafir and Tversky, 1992, 1995).

Although we still do not have adequate models of these heuristics, we do have the following general principles.

1. in judging whether an event or object A belongs to a class or process B, one heuristic people use is to consider whether A is *representative* of B, but not other relevant facts, such as the frequency of B. For instance, if

informed that an individual has a good sense of humor and likes to entertain friends and family, and asked if the individual is a professional comic or a clerical worker, people are more likely to say the former, despite the fact that a randomly chosen person is much more likely to be a clerical worker than a professional comic, and many people have a good sense of humor, so there are many more clerical workers satisfying the description than professional comics.

2. in assessing the frequency of an event, people take excessive accounts of information that is easily *available* or highly *salient*, even though a selective bias is obviously involved. For this reason people tend to overestimate the probability of rare events, since such events are highly newsworthy while nonoccurrences are not reported.

The heuristics and social rules upon which people base their choices and preferences can be expected to reflect their well-being and their evaluation of social welfare only when there is a *feedback loop* from behavior, to social outcomes, to changes in behavior — a dynamic that operates appropriately in our daily lives, but is inoperative in dealing with global, long-term changes in environmental conditions. Other methods than assessing the 'revealed preferences' of citizens and taxpayers in determining environmental benefits and costs are thus worthy of exploration.

2.3. Loss aversion and status quo bias

The *H. economicus* model assumes that people react to the absolute level of payoffs, whereas experiments show that in fact they tend to privilege the *status quo* (their current position) and are sensitive to changes from the *status quo*. Thus experienced well-being is associated more with *changes* in income rather than with the *level* of income — see, for instance, Easterlin (1974, 1995), Lane (1991, 1993) and Oswald (1997).

Experimental evidence supports an even stronger assertion: people tend to exhibit *loss aversion*. Specifically, *people are about twice as averse to taking losses as to enjoying an equal level of gains* (Helson, 1964; Tversky and Kahneman, 1981b; Kahneman et al., 1990). This means, for instance, that an individual may attach zero value to a lottery that offers an equal chance of winning 1000 and losing 500. This also implies that people are *risk-loving over losses*, while they remain risk-averse over gains. For instance, many individuals will choose a 50 probability of losing 2000 rather than losing 1000 with certainty (both have the same expected value, of course, but the former is riskier). The same individuals will choose a certain *gain* of 1000 over a 50% chance of a gain of 2000.

One implication of loss aversion is the *endowment effect* (Kahneman et al., 1991), according to which people place a higher value on what they possess than they place on the same things when they do not possess them. For instance, if you win a bottle of wine that you could sell for $200, you may drink it rather than selling it, but you would never think of buying even a $100 bottle of wine. Not only does the endowment effect exists, but there is evidence that people underestimate it, and hence cannot effectively correct for it in their choice behavior (Loewenstein and Adler, 1995).

Another implication is the existence of a *framing effect*, whereby one form of a lottery is strictly preferred to another, even though they have the same payoffs with the same probabilities (Tversky and Kahneman, 1981a). For instance, people prefer a price of $10 plus a $1 discount to a price of $8 plus a $1 surcharge. Framing is, of course, closely associated with the endowment effect, since framing usually involves privileging the initial state from which movements are assessed.

Yet another implication is a *status quo bias*, according to which individuals often prefer the *status quo* over any of the alternatives, but if one of the alternatives becomes the *status quo*, that too is preferred to any of the alternatives (Kahneman et al., 1991). The *status quo* makes sense if we recognize that any change can involve a loss, and since on the average gains do not offset losses, it is possible that any one of a number of alternatives might be preferred if it is the *status quo*.

If this experimental evidence is generally correct, the costs of reducing economic growth rates (which are reductions in future gains) must be weighted at about half the costs of current environmental destruction (which are reductions in current enjoyments). However, in evaluating the costs of future environment deterioration, future generations will experience a lower level of loss in welfare than would the current generation if obliged to live under the conditions of these future generations.

3. Experiments in strategic interaction

Many experimental games involve not games against nature or our (future) selves, but rather against (or with) others. In this section I will describe the results of some basic experimental games. As a basis for interpreting this broad range of experiments, I will introduce a new *persona* I call *H. reciprocans*. *H. reciprocans*' behavior in market situations, in which punishing and rewarding are impossible or excessively costly, is much like that of *H. economicus*. But *H. reciprocans* comes to strategic interactions with a propensity to cooperate, responds to cooperative behavior by maintaining or increasing his level of cooperation, and responds to noncooperative behavior by retaliating against the 'offenders', even at a personal cost, and even when there is no reasonable expectation that future personal gains will flow from such retaliation. When other forms of punishment are not available, *H. reciprocans* responds to defection with defection, leading to a downward spiral of non-cooperation. *H. reciprocans* is thus neither the selfless altruist of utopian theory, nor the selfish hedonist of neoclassical economics. Rather, *H. reciprocans* is a conditional cooperator whose penchant for reciprocity can be elicited under circumstances in which personal self-interest would dictate otherwise.

3.1. The ultimatum game

In the *ultimatum game*, under conditions of anonymity, one player, called the 'proposer', is handed a sum of money, say $10, and is told to offer any number of dollars, from $1 to $10, to the second player, who is called the 'responder'. The responder, again under conditions of anonymity, can either accept the offer, or reject it. If the responder accepts the offer, the money is shared accordingly. If the responder rejects the offer, both players receive nothing.

According to the *H. economicus* model, the responder will accept any offer, since something is better than nothing, and knowing this, the proposer will offer the minimum possible amount—in our case, 1. However, when actually played by subjects in an experimental setting, *the H. economicus outcome is almost never attained or even approximated*. In fact, as many replications of this experiment have documented, under varying conditions and with varying amounts of money, proposers routinely offer respondents very substantial amounts (50% of the total being the modal offer), and respondents frequently reject offers below 30% (Güth and Tietz, 1990; Roth et al., 1991; Camerer and Thaler, 1995). These results are obtained in experiments with stakes as high as 3 months earnings. For a review of ultimatum game experiments, see Güth and Tietz (1990), Roth (1995), and Camerer and Thaler (1995).

When asked why they offer more than the lowest possible amount, proposers commonly say that they are afraid that respondents will consider low offers unfair and reject them. When respondents reject offers, they give virtually the same reasons for their actions.[4]

The reader may suspect that it is not strong reciprocity, but rather a distaste for unequal divisions (and especially for being on the short side of an unequal division) that induces respondents to reject low offers in the ultimatum game. To test this possibility, Blount (1995) set up an ultimatum game in which a computer rather than the proposer generated the offer. When respondents were told of this fact, rejection rates fell dramatically, suggesting that the motive of punishment of a cooperative norm violation, not simply rejecting an unequal outcome, was at work.[5]

[4] In all of the above experiments a significant fraction of subjects (about a quarter, typically) conform to the self-interested preferences of *H. economicus*, and it is often the self-serving behavior of this minority that, when it goes unpunished, unravels initial generosity and cooperation when the game is repeated.

[5] Experiments also reveal that individuals are somewhat 'inequality averse', exhibiting a weak urge to reduce inequality when on top, and a relatively stronger urge to reduce inequality when on the bottom. For instance, Blount (1995) found that extremely unequal offers were likely to be rejected even when generated by a computer. For a model of this behavior, see Fehr and Schmidt (1999) and Gintis (2000, Ch. 11).

3.2. The public goods game

Another important experimental setting in which strong reciprocity has been observed is that of the *public goods game*, designed to illuminate such problems as the voluntary payment of taxes and contribution to team and community goals. Public goods experiments have been run many times, under varying conditions, beginning with the pioneering work of the sociologist G. Marwell, the psychologist R. Dawes, the political scientist J. Orbell, and the economists R. Isaac and J. Walker in the late 1970s and early 1980s.[6] The following is a common variant of the game. Ten subjects are told that $1 will be deposited in each of their 'private accounts' as a reward for participating in each round of the experiment. For every $1 a subject moves from his 'private account' to the 'public account', the experimenter will deposit $0.50 in the private accounts of each of the subjects at the end of the game. This process will be repeated ten times, and at then end, the subjects can take home whatever they have in their private accounts.

If all ten subjects are perfectly cooperative, each puts $1 in the public account at the end of each round, generating a public pool of $10; the experimenter then puts $5 in the private account of each subject. After ten rounds of this, each subject has $50. Suppose, by contrast, that one subject is perfectly selfish, while the others are cooperative. The selfish player keeps the $1-per-round in his private account, whereas the cooperative players continue to put the $1 in the public pool. In this case, the selfish subject who takes a free ride on the cooperative contributions of others ends up with $55 at the end of the game, while the other players end up with $45 each. But if all players opt for the selfish payoff, then no one contributes to the public pool, and each ends up with $10 at the end of the game. And if one player cooperates, while the others are all selfish, that player will end up with $5 at the end of the game, while the others will get $15. It is thus clear that this is indeed an 'iterated prisoner's dilemma' — whatever other players do on a particular round a player's highest payoff comes from contributing nothing to the public account. If others cooperate, it is best to take a free ride; if others are selfish, it is best to join them. But if no one contributes, all receive less than they would had all cooperated.

Public goods experiments find that only a fraction of subjects conform to the *H. economicus* model, contributing nothing to the public account. Rather, in a one-stage public goods game, people contribute on average about half of their private account. The results in the early stages of a repeated public goods game are similar. In the middle stages of the repeated game, however, contributions begin to decay until at the end, they are close to the *H. economicus* level — i.e. zero.

Could we not explain the decay of public contribution by *learning*: the participants really do not understand the game at first, but once they hit upon the free-riding strategy, they apply it? Not at all. One indication that learning does not account for the decay of cooperation is that increasing the number of rounds of play (when this is known to the players) leads to a decline in the rate of decay of cooperation (Isaac et al., 1994). Similarly, Andreoni (1988) finds that when the whole process is repeated with the same subjects, the initial levels of cooperation are restored, but once again cooperation decays as the game progresses. Andreoni (1995) suggests a *H. reciprocans* explanation for the decay of cooperation: public-spirited contributors want to retaliate against free-riders and the only way available to them in the game is by not contributing themselves.

3.3. The public goods game with retaliation

Could the decay of cooperation in the public goods game be due to cooperators retaliating against free-riders by free-riding themselves? Subjects often report this behavior retrospectively. More compelling, however, is the fact that when subjects are given a more constructive way of punishing defectors, they use it in a way that helps sustain cooperation (Dawes et al., 1986; Sato 1987; Yamagishi 1988a,b, 1992).

For instance, in Ostrom et al. (1992) subjects interacted for about 25 periods in a public goods

[6] For a summary of this research and an extensive bibliography, see Ledyard (1995).

game, and by paying a 'fee', subjects could impose costs on other subjects by 'fining' them. Since fining costs the individual who uses it, but the benefits of increased compliance accrue to the group as a whole, no self-interested *H. economicus* will pay the fee, so no player is ever punished for defecting, and all self-interested players defect by contributing nothing to the public account. However, the authors found a significant level of punishing behavior. The experiment was then repeated with subjects being allowed to communicate, but without being able to make binding agreements. In the framework of the *H. economicus* model, such communication is called *cheap talk*, and cannot lead to cooperation, since there is no cost to reneging on a promise. But in fact such communication led to almost perfect cooperation (93) with very little sanctioning (4).

The design of the Ostrom et al. study allowed individuals to engage in strategic behavior, since costly retaliation against defectors could increase cooperation in future periods, yielding a positive net return for the retaliator. It is true that logical thinking rules out such a strategy for the 'rational agents' of traditional economics, but we know that individuals are in fact not 'rational' in this sense — in fact they are more rational, in that when subjects interact in finitely repeated games, they do better than the 'rational' agents of neoclassical economics, as reported in McKelvey and Palfrey (1992) and Gintis (2000, Ch. 2). What happens if we remove *any possibility* of retaliation being strategic? This is exactly what Fehr and Gächter (2000) studied. They set up a repeated public goods game with the possibility of costly retaliation, but they ensured that group composition changed *in every period* so subjects knew that costly retaliation could not confer any pecuniary benefit to those who punish. Nonetheless, punishment of free-riding was prevalent and gave rise to a large and sustainable increase in cooperation levels. Indeed, even on the last round of the game cooperation remains at a very high level.

3.4. The common pool resource game

In 1968 Garrett Hardin wrote a famous article in the journal *Science* entitled 'The Tragedy of the Commons' (Hardin, 1968). The term 'commons' referred originally to the region of an English village that belonged to the villagers as a group, and on which villagers were permitted to graze their sheep or cows. The 'tragedy' in the tragedy of the commons was that the commons tended to be overgrazed, since each villager would graze to the point where the *private* costs equals the benefits, whereas grazing imposed additional *social* costs on the rest of the community. This situation applies to what are termed *common pool resources* in general. Some involve social problems of the highest environmental importance, including air and water pollution, overfishing, overuse of antibiotics, excessive groundwater use, and overpopulation.

The general implication from Hardin's analysis was that some centralized entity, such as a national government or international agency, would have to step in to prevent the tragedy by regulating the common. The historical experience in regulating the commons was, however, a patchwork of successes and failures. Elinor Ostrom (1990) published an influential book, *Governing the Commons*, suggesting that the Hardin analysis did not apply generally, since local communities often had ways of self-organizing and self-governing to prevent overexploitation of the commons, and that government policy often exacerbated rather than ameliorated the problem by undermining the social connections on which local regulation was based.

When formalized as a game, the common pool resource problem is simply an n-person repeated prisoner's dilemma, in which each player hopes the other players will cooperate (not take too much of the common resource), but a player who acts like *H. economicus* will defect (take too much) no matter what the others do. In fact, both in real world and experimental settings, under the appropriate conditions we see much more cooperation than predicted by the *H. economicus* model.

Ostrom et al. (1994) used both experimental and field data to test game-theoretic models of common pool resources. They found more spontaneous cooperation in the field studies than predicted, and when communication and sanctioning were permitted in the laboratory, the level of cooperation became quite high.

While common pool resource and public goods games are equivalent for *H. economicus*, people treat them quite differently in practice. This is because the *status quo* in the public goods game is the individual keeping all the money in the private account, while the *status quo* in the common pool resource game the resource not being used at all. This is a good example of a *framing effect*, since people measure movements from the *status quo*, and hence tend to undercontribute in the public goods game, and overcontribute (underexploit) in the common pool resource game, compared to the social optimum (Ostrom, 1998).

It is clear that in the real world, of course, communities often do *not* manage their common pool resources well. The point of Ostrom's work is to identify the sources of failure, not to romanticize small communities and informal organization. The management of common pool resources fails, for instance, when communities are so large that it pays to form a local coalition operating against the whole community, or when resources are so unequally distributed that it pays the wealthy to defect on the nonwealthy and conversely (Hackett et al., 1994; Bardhan et al., 1999).

4. Interpreting experimental results

How are we to interpret the various empirical results documented above? When the results of experiments contradict received wisdom in economics, many economists reject the experiments rather than received wisdom. For instance, in the ultimatum game, individuals frequently choose zero payoffs when positive payoffs are available. Critics claim that subjects have not learned how to play the game and are confused by the unreality of the experimental conditions, so their behavior does not reflect real life. Moreover whatever experimentalists do to improve laboratory protocols (e.g. remove cues and decontextualize situations) the critics deem as insufficient, and the experimentalists complain among themselves that the critics are simply dogmatic enemies of the 'scientific method'.

To move beyond this impasse we must recognize that the critics are correct in sensing some fundamental difference between experiments in social interaction and the traditional experimental method in natural science, and that experimental results must be interpreted more subtly than is usually done. The upshot is, however, an even stronger vindication of the experimental method, and an even deeper challenge to the received wisdom.

Laboratory experiments are a means of controlling the social environment so that experiments can be replicated and the results from different experiments can be compared. In physics and chemistry the experimental method has the additional goal of *eliminating all influences on the behavior of the object of study except those controlled by the experimenter*. This goal can be achieved because elementary particles, and even chemical compounds, are completely interchangeable, given a few easily measurable characteristics (atomic number, energy, spin, chemical composition, and the like). Experiments in human social interaction, however, *cannot* achieve this goal, even in principle, because experimental subjects bring their personal history with them into the laboratory. Their behavior is therefore *ineluctably* an interaction between the subject's personal history and the experimenter's controlled laboratory conditions.

Neither personal history nor general cultural/genetic evolutionary history has prepared subjects for the ultimatum, public goods, common pool resource, and other games that we shall describe below. In fact, an agent reacts to a totally novel situation by first assigning the situation to one of a small number of pre-given *situational contexts*, and then deploying the behavioral repertoire — payoffs, probabilities, and actions — appropriate to that context. We may call this *choosing a frame* for interpreting the experimental situation.

The results of the ultimatum game, for instance, suggest that in a two-person bargaining situation, in the absence of other cues, the situational context applied by the subject involves 'sharing'. Suppose we change the rules such that both proposer and respondent are members of different *teams* and each is told that their respective winnings will be paid to the team rather than the individual. A distinct situational context involving 'winning' is now often deemed appropriate, dictating acting on behalf of one's team and suppressing behaviors that would be otherwise individually satisfying—such as 'sharing'. In this case proposers offer much less, and respondents very rarely reject positive offers (Shogren, 1989). Similarly, if the experimenters introduce notions of property rights into the strategic situation (e.g. that the proposer in an ultimatum game has 'earned' or 'won' the right to this position), then motivations concerning 'fairness' are considerably attenuated in the experimental results (Hoffman et al., 1994, 1996).

In short, laboratory experiments elucidate how subjects identify situational contexts and then describe how agents react to the formal parameters and material payoffs, subject to the situational contexts they have identified.

5. Conclusion: self-interest and rationality

The culture surrounding economics as a discipline fosters the belief that rationality implies self-interest, outcome-orientation, and time-consistency. If this were correct, we would have to call real-life humans hopelessly irrational. The economist's treatment of rationality, however, cannot be supported.

A *rational agent* is one who draws conclusions logically from given premises, whose premises are defensible by reasoned argument, who uses evidence dispassionately in evaluating factual assertions, and more technically, who optimizes subject to constraints under conditions of limited information and costly decision-making.

I have never encountered a credible argument supporting the assertion that rationality in this sense implies self-interest, outcome-orientation, or time-consistency. In particular, it is just as 'rational' for me to prefer to have *you* enjoy a fine meal as for me to enjoy the meal myself. It is just as 'rational' for me to care about the rain forests as to care about my beautiful cashmere sweater. And it is just as 'rational' to reject an unfair offer as it is to discard an ugly article of clothing. Supporters of the traditional model of self-interested actors often suggest that the process of Darwinian evolution will weed out altruistic and other non-self-interested actors, but there are general arguments (see Sober and Wilson (1998) and references therein) and analytical models (see Gintis (2000), Ch. 11 and references therein) that show that this need not be the case.

It is clear that economic theory has much to offer in formulating principles of environmental regulation and in evaluating environmental policies, but its contributions will be considerably more valuable when *H. economicus* is replaced by a more accurate model of individual choice and strategic interaction.

Acknowledgements

I would like to thank the MacArthur Foundation for research support.

References

Ainslie, G., 1975. Specious reward: a behavioral theory of impulsiveness and impulse control. Psychol. Bull. 82, 463–496.

Ainslie, G., Haslam, N., 1992. Hyperbolic discounting. In: Loewenstein, G., Elster, J. (Eds.), Choice Over Time. Russell Sage, New York, pp. 57–92.

Andreoni, J., 1988. Why free ride? strategies and learning in public good experiments. J. Pub. Econ. 37, 291–304.

Andreoni, J., 1995. Cooperation in public goods experiments: kindness or confusion. Am. Econ. Rev. 85 (4), 891–904.

Anscombe, F., Aumann, R., 1963. A definition of subjective probability. Annals Math. Stat. 34, 199–205.

Axelrod, R., 1984. The Evolution of Cooperation. Basic Books, New York.

Bardhan, P., Bowles, S., Gintis, H., 1999. Wealth inequality, credit constraints, and economic performance. In: Atkinson, A., Bourguignon, F. (Eds.), Handbook of Income Distribution. North-Holland, Dordrecht.

Blount, S., 1995. When social outcomes aren't fair: the effect of causal attributions on preferences. Organ. Behav. Hum. Decis. Process. 63(2), 131–144.

Camerer, C., Thaler, R., 1995. Ultimatums, dictators, and manners. J. Econ. Perspect. 9, 209–219.

Chung, S.-H., Herrnstein, R.J., 1967. Choice and delay of reinforcement. J. Exp. Anal. Behav. 10, 67–74.

Dawes, R.M., Orbell, J.M., Van de Kragt, J.C., 1986. Organizing groups for collective action. Am. Pol. Sci. Rev. 80, 1171–1185.

Easterlin, R.A., 1974. Does Economic Growth Improve the Human Lot? Some Empirical Evidence. In: Nations and Households in Economic Growth: Essays in Honor of Moses Abramovitz, Academic Press, New York.

Easterlin, R.A., 1995. Will raising the incomes of all increase the happiness of all? J. Econ. Behav. Organ. 27, 35–47.

Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition, and cooperation. Quart. J. Econ. 114, 817–868.

Fehr, E., Zych, P., 1994. The Power of Temptation: Irrationally Myopic Excess Consumption in an Addiction Experiment, Unpublished Manuscript, University of Zurich.

Fehr, E., Gächter, S., 2000. Cooperation and Punishment, American Economic Review, forthcoming.

Friedman, M., Savage, L.J., 1948. The utility analysis of choices involving risk. J. Political Econ. 56, 279–304.

Gintis, H., 2000. Game Theory Evolving. Princeton University Press, NJ.

Güth, W., Tietz, R., 1990. Ultimatum bargaining behavior: a survey and comparison of experimental results. J. Econ. Psychol. 11, 417–449.

Hackett, S., Schlager, E., Walker, J., 1994. The role of communication in resolving commons dilemmas: experimental evidence with heterogeneous appropriators. J. Environ. Econ. Manag. 27, 99–126.

Hardin, G., 1968. The tragedy of the commons. Science 162, 1243–1248.

Helson, H., 1964. Adaptation Level Theory: An Experimental and Systematic Approach to Behavior. Harper and Row, New York.

Herrnstein, R.J., Prelec, D., 1992. A theory of addiction. In: Loewenstein, G., Elster, J. (Eds.), Choice over Time. Russell, Sage, New York, pp. 331–360.

Hoffman, E., McCabe, K., Smith, V.L., 1996. Social distance and other-regarding behavior in dictator games. Am. Econ. Rev. 86, 653–660.

Hoffman, E., McCabe, K., Shachat, K., Smith, V.L., 1994. Preferences, property rights, and anonymity in bargaining games. Games Econ. Behav. 7, 346–380.

Isaac, R.M., Walker, J.M., Williams, A.W., 1994. Group size and voluntary provision of public goods: experimental evidence utilizing large groups. J. Public Econ. 54, 1–36.

Kahneman, D., Knetch, J.L., Thaler, R.H., 1990. Experimental tests of the endowment effect and the Coase theorem. J. Pol. Econ. 98 (6), 1325–1348.

Kahneman, D., Knetch, J.L., Thaler, R.H., 1991. The endowment effect, loss aversion, and status quo bias. J. Econ. Persp. 5 (1), 193–206.

Kahneman, D., Slovic, P., Tversky, A., 1982. Judgment under Uncertainty: Heuristics and Biases. Cambridge University Press, Cambridge, UK.

Kirby, K.N., Herrnstein, R.J., 1995. Preference reversals due to myopic discounting of delayed reward. Psychol. Sci. 6 (2), 83–89.

Laibson, D., 1997. Golden eggs and hyperbolic discounting. Quart. J. Econ. 112 (2), 443–477.

Laibson, D., Rachlin, H., 1997. The Matching Law: Papers on Psychology and Economics by Richard Herrnstein, Harvard University Press, Cambridge, MA.

Lane, R., 1991. The Market Experience. Cambridge University Press, Cambridge, UK.

Lane, R.E., 1993. Does money buy happiness? The Public Interest 113, 56–65.

Ledyard, J.O., 1995. Public Goods: A Survey of Experimental Research. In: Kagel, J.H., Roth, A.E. (Eds.), The Handbook of Experimental Economics. Princeton University Press, Princeton, NJ, pp. 111–194.

Loewenstein, G., 1987. Anticipation and the valuation of delayed consumption. Econ. J. 97, 666–684.

Loewenstein, G., Adler, D., 1995. A bias in the prediction of tastes. Econ. J. 105 (431), 929–937.

Loewenstein, G., Prelec, D., 1992. Anomalies in intertemporal choice: evidence and an interpretation. Quart. J. Econ. 57, 573–598.

Loewenstein, G., Sicherman, N., 1991. Do workers prefer increasing wage profiles? J. Labor Econ. 91 (1), 67–84.

McKelvey, R.D., Palfrey, T.R., 1992. An experimental study of the centipede game. Econometrica 60, 803–836.

Ostrom, E., 1990. Governing the Commons: The Evolution of Institutions for Collective Action. Cambridge University Press, Cambridge, UK.

Ostrom, E., 1998. A behavioral approach to the rational choice theory of collective action. Am. Political Sci. Rev. 92 (1), 1–21.

Ostrom, E., Walker, J., Gardner, R., 1992. Covenants with and without a sword: self-governance is possible. Am. Political Sci. Rev. 86 (2), 404–417.

Ostrom, E., Gardner, R., Walker, J., 1994. Rules, Games, and Common-Pool Resources. University of Michigan Press, Ann Arbor.

Oswald, A.J., 1997. Happiness and economic performance. Econ. J. 107 (445), 1815–1831.

Robbins, L., 1935. An Essay on the Nature and Significance of Economic Science. Macmillan, London.

Roth, A., 1995. Bargaining experiments. In: Kagel, J., Roth, A. (Eds.), The Handbook of Experimental Economics. Princeton University Press, Princeton, NJ.

Roth, A.E., Prasnikar, V., Okuno-Fujiwara, M., Zamir, S., 1991. Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: an experimental study. Am. Econ. Rev. 81 (5), 1068–1095.

Sato, K., 1987. Distribution and the cost of maintaining common property resources. J. Exp. Soc. Psychol. 23, 19–31.

Savage, L.J., 1954. The Foundations of Statistics. Wiley, New York.

Shafir, E., Tversky, A., 1992. Thinking through uncertainty: nonconsequential reasoning and choice. Cogn. Psychol. 24 (4), 449–474.

Shafir, E., Tversky, A., 1995. Decision making. In: Smith, E.E., Osherson, D.N. (Eds.), Thinking: An Invitation to Cognitive Science, vol. 3. MIT Press, Cambridge, MA, pp. 77–100.

Shogren, J.F., 1989. Fairness in bargaining requires a context: an experimental examination of loyalty. Econ. Lett. 31, 319–323.

Sober, E., Wilson, D.S., 1998. Unto Others: The Evolution and Psychology of Unselfish Behavior. Harvard University Press, Cambridge, MA.

Trivers, R.L., 1971. The evolution of reciprocal altruism. Quart. Rev. Biol. 46, 35–57.

Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: heuristics and biases. Science 185, 1124–1131.

Tversky, A., Kahneman, D., 1981a. The framing of decisions and the psychology of choices. Science 211, 453–458.

Tversky, A., Kahneman, D., 1981b. Loss aversion in riskless choice: a reference-dependent model. Quart. J. Econ. 106, 1039–1061.

Von Neumann, J., Morgenstern, O., 1944. Theory of Games and Economic Behavior. Princeton University Press, Princeton, NJ.

Yamagishi, T., 1988a. The provision of a sanctioning system in the United States and Japan. Soc. Psychol. Quart. 51 (3), 265–271.

Yamagishi, T., 1988b. Seriousness of social dilemmas and the provision of a sanctioning system. Soc. Psychol. Quart. 51 (1), 32–42.

Yamagishi, T., 1992. Group size and the provision of a sanctioning system in a social dilemma. In: Liebrand, W.B.G., Messick, D.M., Wilke, H.A.M. (Eds.), Social Dilemmas: Theoretical Issues and Research Findings. Pergamon Press, Oxford, pp. 267–287.