Economic Man —or Straw Man?

A Commentary on Henrich *et al*

Ken Binmore
Economics Department
University College London
Gower Street
London WC1E 6BT, UK

Economic Man—or Straw Man?

A Commentary on Henrich *et al*

by Ken Binmore

1 Ignoratio Elenchi

This commentary on the paper *"Economic Man" in Cross-Cultural Perspective* [20] is fiercely critical, but the criticism is not directed at the anthropological field work reported in the paper, which seems to me entirely admirable.

The criticism is directed at the editorial rhetoric that accompanies the scientific reports of the experiments carried out in the fifteen small-scale societies studied. The rhetoric is markedly more subdued than in the book *Foundations of Human Sociality* [19] from which the current paper is extracted. (See Samuelson [27] for a review.) However, the claim remains that "economic man" is an experimental failure, and that we must seek an alternative paradigm.

This paper argues that the editors' enthusiasm for this perennially popular claim has led them into two mistakes. Philosophers call the first mistake the *ignoratio elenchi*—the refuting of propositions that your opponent does not maintain. In particular, it is not axiomatic in orthodox economic theory that human beings are selfish. Even if such a proposition were axiomatic, the backward induction principle the authors use when analyzing the Ultimatum Game would not follow.

The second mistake is that of neglecting to report data that does not support their claims about "economic man". In particular, although it is not axiomatic in mainstream economics that human beings maximize their own income, there is a huge experimental literature whose results are consistent with the hypothesis that most people behave in this way after gaining sufficient experience of most tasks they are set in the laboratory.

As a result of these mistakes, the editors contrive to treat conclusions of their study that are broadly supportive of the game-theoretic approach to social norms as though they were inconsistent with the principles on which game theory is based.

2 De Gustibus Non Est Disputandum

It is not true that "texbook predictions" based on *Homo economicus* incorporate a "selfishness axiom". The orthodox position amongst economists is embodied in Paul Samuelson's theory of revealed preference, which makes a virtue of refusing to make any *a priori* hypotheses at all about what goes on inside people's heads.

The orthodox theory only requires that people behave consistently. It is then shown that they will then necessarily behave *as though* they are maximizing something. Economists call this something *utility*, but they emphatically do not argue that people have little utility generators in their heads. Still less do they argue that people come equipped with mental cash registers that respond only to dollars.

Far from making it axiomatic that human beings maximize money, the orthodoxy is that the nature of a person's utility function must be determined *empirically* by observing his or her choice behavior in some situations. If the person behaves consistently, the utility function then serves as a tool in predicting how the person will choose in other situations.

3 Empirical Data on Selfish Behavior

We have seen that Henrich *et al* [20] are mistaken in suggesting that it is *axiomatic* in mainstream economics that people behave selfishly. Orthodox economic theory leaves the question of how selfish real people actually are to be decided by empirical means. However, Henrich *et al* are right that mainstream economists do commonly think that we must expect to see a lot of selfishness from experienced subjects in laboratory experiments. If they want to challenge this consensus, they need to address the *empirical* evidence that mainstream economists offer in its support.

Henrich *et al* [20] tell us that "hundreds of experiments" show that subjects will "sacrifice their own gains to change the distribution of material outcomes among others". We are not told that for every experiment whose results the authors choose to interpret in this way, there are large numbers of other experiments in which orthodox game theory predicts the outcome reasonably well on the naive assumption that subjects who are sufficiently experienced behave as though simply seeking to maximize their average payoff in money.

In neglecting to draw attention to this empirical fact, Henrich *et al* might feel that they are entitled to ignore the experimental work of mainstream experimental economists—like Vernon Smith and Charles Plott—on the grounds that market institutions tend to make people behave selfishly. But even if experiments about markets are omitted, there remains a huge literature in which subjects are reported as learning to play like income maximizers in games that are not structured like markets. This fact is now so trite that it is hard to get a paper published if its only finding is that the subjects' behavior eventually converges on a Nash equilibrium in yet another game with money payoffs. One might conceivably argue that the authors are entitled to regard such work—including my own—as dubious, because it is carried out by economists who are supposedly biased in favor of a "selfishness axiom" (Binmore [8, 9]). But I do not see how it is possible to justify passing over the empirical evidence in the case of one of the games in their own experimental repetoire.

The Prisoners' Dilemma is the most famous example of a Public Goods game. The essence of such games is that each player can privately make a contribution to

a notional public good. The sum of contributions is then increased by a substantial amount and the result redistributed to all the players. In such games, it is optimal for a selfish player to "free ride" by contributing nothing, thereby pocketing his share of the benefit provided by the contributions of the other players without making any contribution himself.

We are told that students in such Public Goods games contribute a mean amount of between 40% and 60% of the total possible, but that this "fairly robust" conclusion is "sensitive to the costs of cooperation and repeated play". But we are not told *how* sensitive. In fact, the results are very sensitive indeed to repeated play. If their remarks are intended to apply to all the games in their experimental repetoire, then Henrich *et al* [20] stand the truth on its head when they claim that:

> Initial skepticism about such experimental evidence has waned as subsequent studies involving high stakes and ample opportunity for learning has repeatedly failed to modify these fundamental conclusions.

This claim that learning can be disregarded is all the more remarkable in that some of the authors—notably Camerer (author number four) [12, p.265]—have themselves published experimental papers which confirm that it can often matter very much indeed. In the case of Public Goods games, the standard result is exemplified by the first ten trials of an experiment of Fehr (author number five) and Gächter [14] illustrated in Figure 3.2 of Henrich *at al* [19].

Camerer explicitly endorses this result as standard in his recent *Behavioral Game Theory* [12, p.46]. As he says, the huge number of experimental studies available in 1995 was surveyed both by John Ledyard [22] and by David Sally [24], the former for Roth and Kagel's authoritative *Handbook of Experimental Economics.* After playing repeatedly (against a new opponent each time), the much replicated result is that about 90% of subjects end up free riding. One can disrupt the march towards free riding in various ways, but when active intervention ceases, the march resumes.

I emphasize these conclusions because the orthodox view among mainstream economists and game theorists who take an interest in experimental results is not that the *learning* that might take place during repeated play in the laboratory is a secondary phenomenon to which conclusions may or may not be sensitive. On the contrary, the fact that laboratory subjects commonly adapt their behavior to the game they are playing as they gain experience is entirely central to our position. There is therefore no point in our critics endlessly refuting the proposition that inexperienced or underpaid subjects in novel situations do not play as game theory predicts. Everybody accepts that this is true most of the time.

Before leaving this subject, let me observe that nobody maintains that *all* subjects learn to behave like selfish optimizers in *all* environments. The 10% or so of subjects who choose to continue cooperating in Public Goods games in spite of their experience are a small group, but nobody denies their existence. As for an environment in which subjects do not seem to adapt their behavior as they gain experience, we have the Ultimatum Game, discussed below in Section 6.

Nor does mainstream empirical work support the conclusion that most expe-

rienced people have no other-regarding or social component at all in the utility functions that describe their final choices. It only supports the conclusion that for most adequately incentified people in most economic environments in developed societies, the data can be explained without assuming that such an other-regarding component is large. However, one would not expect the same conclusion in societies where kinship is a major social factor.

Nor do mainstream economic theorists hold that the small deviations from selfish optimization that may well occur in situations in which the hypothesis of income maximization predicts the data well are necessarily unimportant in other situations. On the contrary, one of the advantages of a game-theoretic approach is that it is capable of predicting that such small deviations are likely to have large effects in certain sensitive games—like the finitely repeated Prisoners' Dilemma mentioned in Section 4, or the Public Goods Game with Punishment that is the subject of the second ten trials in Figure 3.2 of Henrich *et al* [20].

4 Social Norms as Equilibria in Games

Up to now, this commentary has focused on what orthodox economic theory does not maintain, but one may reasonably ask what positive insights orthodox economic theorists think they may be able to contribute to anthropological studies of small-scale traditional societies.

Our view is that the simplest game that can reasonably model life in such a community is an indefinitely *repeated* game in which the players always remain the same. There is a simple theorem called the *folk theorem* that applies to those indefinitely repeated games in which the players have no secrets from each other and care a lot about their own future welfare.[1] The folk theorem says that such repeated games always have a large number of Nash equilibria—which is the concept that game theorists use to predict the behavior of experienced players. The theorem says that *any* outcome on which the players might want to contract in the presence of an external agency ready and willing to enforce contracts is available as an equilibrium outcome if the players are sufficiently forward-looking.

The significance of this result is that Nash equilibrium outcomes do not need external enforcement to be viable. They are *self-enforcing,* since nobody can gain by deviating from a Nash equilibrium unless someone else deviates first.

Axelrod's [2] widely cited *Evolution of Cooperation* is misleading on the subject of repeated games, not least because his emphasis on the tit-for-tat strategy distracts attention from the important fact that there are always *many* Nash equilibria that support efficent cooperation among the players. Like the tit-for-tat strategy, all such equilibrium strategies require behavior that would normally be described as

[1] The theorem is called the folk theorem, because nobody knew to whom it should be attributed when it was discovered simultaneously by a number of game theorists in the early 1950s. Aspects of the folk theorem have been repeatedly rediscovered, notably by Boyd (author number two) and Richerson [11] in 1992, Axelrod [2] in 1984, and Trivers [30] in 1971.

embodying a strong respect for both *reciprocity* and *reputation*—even though no concern for such emergent phenomena need be built into the utility functions assigned to the players. Experimental findings that reciprocity and reputation matter in some social interactions therefore come as no surprise to orthodox game theorists.

The biologist Robert Trivers [30] says that strategies which support full cooperation in repeated games exhibit *reciprocal altruism.* He cares about this phenomenon because Nash equilibria do not only describe the outcome of rational play; under appropriate conditions, they also describe the end-product of evolutionary processes. It is therefore of some importance that the only one of Axelrod's conclusions that seems to be genuinely robust is his claim that we should normally expect the kind of evolutionary computer simulations that he pioneered to lead to efficient (fully cooperative) Nash equilibria (Binmore [4, p.313]).

A social norm can be seen as a device for solving the equilibrium selection problem that the folk theorem says is built into a society's indefinitely repeated "game of life". We then obtain a putative explanation for the cultural evolution of different social norms in different societies. The folk theorem therefore provides a theoretical backdrop for the ideas on cultural evolution pioneered by Boyd (author number two) and Richerson [10, 11].

The important point is that only the Nash equilibria of a society's game of life can be evolutionarily stable, but nothing says that evolution must select the *same* equilibrium in different societies. (See Samuelson [26] and Young [31].) Just as the French use the Nash equilibrium for the Driving Game in which everyone drives on the right and the English use the Nash equilibrium in which everyone drives on the left, so we must expect different social norms that select different equilibria to have evolved in different societies.

The social norms that interest me most are those that we normally describe in terms of fairness or justice (Binmore [3, 4]). My recent *Natural Justice* [5] offers an algebra-free version of my theory of fairness that draws on both psychological and anthropological thinking. In passing, I try to explain that the reason neo-conservative economists see no role for fairness in their models of the world is that their absurdly over-simplified models only have one equilibrium, and so there is no equilibrium selection problem for fairness to solve. I do not like the policies advocated by such economists any more than the authors of the paper we are discussing, but the answer is not to throw the baby out with the bathwater by seeking to discredit the basic methodology of economics using whatever rhetoric seems currently persuasive, but to direct the attention of neo-conservatives toward economic models that are genuinely descriptive.

Game theorists think the folk theorem is particularly relevant to the social norms (or social contracts) of small-scale societies, because—unlike our own large societies—the no-secrets proviso of the folk theorem has a good chance of being reasonably descriptive. It is no problem that kinship is likely to be an important explanatory variable in such societies, because we can simply write this fact into a player's utility function using an appropriate version of Hamilton's rule before appealing to the folk theorem. (Remember that the idea that our theories necessarily

depend on a "selfishness axiom" is a canard.)

If this attempt to apply game theory in an anthropological context has some descriptive validity, what should we expect to happen when we ask inexperienced subjects from small-scale societies to participate in a novel laboratory game designed to provide information on how people respond to situations involving social phenomena like fairness, trust, or reciprocity? The answer that seems obvious to me is that we should expect them to behave as they would behave in real life if they were offered similar cues to those offered in the laboratory. That is to say, we should use whichever equilibrium their own society operates in its *repeated* game of life to predict their intitial behavior, rather than one of the equilibria of the *one-shot* game they are required to play in the laboratory.

And this seems to be broadly what happens. As Henrich *et al* [20] say: "Experimental play often reflects patterns of interaction found in everyday life." The anthropologist, Jean Ensminger (author number ten), is more explicit when commenting on why the Orma contributed generously in her Public Goods Game:

> When this game was first described to my research assistants, they immediately identified it as the '*harambee*' game, a Swahili word for the institution of village-level contributions for public goods projects such as building a school. ... I suggest that the Orma were more willing to trust their fellow villagers not to free ride in the Public Goods Game because they associated it with a learned and predictable institution. While the game had no punishment for free-riding associated with it, the analogous institution with which they are familiar does. A social norm had been established over the years with strict enforcement that mandates what to do in an exactly analogous situation. It is possible that this institution 'cued' a particular behavior in this game (Henrich *at al* [19, p.376]).

The enforcement here is enforcement by the players themselves as envisaged in the folk theorem, and not external enforcement by the government. (National or cross-regional attempts at *harambee* collections are predictably corrupt.)

If Ensminger is right, then it would be a huge mistake to try to explain the behavior of the Orma in the Public Goods Game on the hypothesis that their behavior was adapted to the game they played in her makeshift laboratory. In particular, inventing other-regarding utility functions whose maximization would lead to generous contribution in the Public Good Game would be pointless. Ensminger is suggesting that the subjects' behavior is adapted to the Public Goods game embedded in the *repeated* game that they play every day of their lives, for which the folk theorem provides an explanation that does not require us to invent anything at all.

Of course, if the subjects play a laboratory game repeatedly (against a new opponent each time), then mainstream economic theory predicts that their behavior would eventually diverge from the equilibrium of the repeated game they are accustomed to play in real life to some equilibrium of the one-shot game they are actually playing in the laboratory.[2] As observed in Section 2, contrary to the impression given

[2]There is a risk of confusion when the repeated play of a one-shot game is under discussion. The assumption is then that players never expect to interact with their current opponents again.

by Henrich *et al* [20], such adaptation to the strategic realities of the actual game being played is uncontroversially the norm in most economic experiments carried out with western undergraduates.

To what extent does such trial-and-error learning occur in the societies studied in *"Economic Man" in Cross-Cultural Perspective?* I do not know the answer because no data on this subject is reported, either in the paper or in the book from which it was extracted. My guess is that in most—perhaps all—of the experiments the subjects never played the same game twice, as is clearly the case in Ensminger's account of her experiment. If so, then these experiments did not offer "ample time for learning". They offered no opportunity for learning at all.

The results of the experiments are therefore incomparable with the mainstream experimental-economics literature in which income-maximizing behavior is commonly reported only after the subjects have had the opportunity to learn about each other and the game they are playing. But the rhetoric adopted by Henrich *et al* [20] obscures this point by pretending that no account need be taken of learning when discussing the data on inexperienced subjects to which they choose to restrict their attention.

However, far from their findings flying in the face of orthodox wisdom, they seem rather to constitute an endorsement of the game-theoretic approach. Game theorists are not in the least surprised to find that the data supports the view that:

> As a consequence of these adaptive learning processes, societies with different historical trajectories are likely to arrive at different social equilibria. Henrich *et al* [20]

But game theorists go further by predicting that when the game of life being played by a society changes, then its social norms will also eventually change by ceasing to coordinate behavior on an equilibrium of the old game and coordinating behavior instead on an equilibrium of the new game. When the anthropological authors of *"Economic Man" in Cross-Cultural Perspective* write on their own behalf, they seem to agree. For example, Michael Alvard (author number eight) tells us that:

> As the results in this volume show, people do not universally play fair. The question is no longer why people seem to have a preference for fairness. The question is now: do people behave more or less fairly in adaptive ways? Henrich *et al* [19, p.433]

We differ only in never having thought that the way to model fairness is to write a taste for fairness into the utility functions of the players.[3]

It is interesting that the change in the economic means of production recently forced on the Ache of Paraguay provides a natural experiment on this issue. Our

---

Unlike the repeated games to which the folk theorem applies, selfish optimizers will then have no reason to take account of either reciprocity or reputation.

[3]Ernst Fehr (author number five) has been prominent in offering experimental evidence to the contrary, but we do not accept that his claims are supported by the data. Shaked [29] provides documentation in the case of Fehr and Schmidt's [15] theory of inequity aversion. Fehr and Schmidt's [16] reply to Shaked's intemporate note makes it clear that they are indeed guilty of what seem to me the more telling of Shaked's accusations of unscientific practice.

prediction would be that the Ache would begin by trying to operate on the basis of the same social contract that they operated as hunter-gatherers—which might be idealistically described by saying that each contributes according to his ability and receives according to his need. But the result of applying such a social contract is not likely to be an equilibrium of their new farming game, because one would expect the old social norms to be destabilized by the emergence of free riders. Cultural evolution should then be expected eventually to generate new social norms that do succeed in coordinating the players' behavior on an equilibrium of the new game. Judging from experience elsewhere, these new social norms would incorporate a stronger sense of private property and less social sharing.

It is hard to estimate the extent to which Ache social norms are adapting to their new game of life from Hill (author number twelve) and Gurven's (author number fourteen) [19, p.338] account, but mainstream economists would be surprised if the Ache do not eventually adapt to their new environment in much the same way as western undergraduates adapt to the one-shot Prisoners' Dilemma.

5 Backward Induction

This section returns to the theme that the economic man of *"Economic Man" in Cross-Cultural Perspective* is made of straw. The authors proceed as though their "selfishness axiom" predicts that only *subgame-perfect* Nash equilibria will be observed. That is to say, that players will solve games by backward induction, as described for the Ultimatum Game in their paper. The same line is taken on "income maximizing" in Henrich *et al* [19].

It is hard to believe that the economists on the panel of authors do not know that this claim is at best controversial. To defend backward induction, one needs not only that it is common knowledge among the players that they are all utility maximizers, but that they disregard any evidence to the contrary that they might receive when playing the game.

To see how unreasonably strong such assumptions are, consider the game obtained by repeating the Prisoners' Dilemma 100 times. The only subgame-perfect equilibrium of this (finitely repeated) game requires that both players always plan to defect at every repetition. But this is not what is observed in the laboratory. Most experienced subjects cooperate until the closing stages of the game, at which point they try to take advantage of their opponent by being the first to defect (Selten *et al* [28]). However, nobody thinks that such results refute an income-maximizing hypothesis, because of the famous gang-of-four paper, which shows that behavior like that observed in the laboratory would be optimal in the presence of a tiny fraction of "irrational" players who always play the tit-for-tat strategy. (See Kreps, Milgrom, Roberts and Wilson [21].)

The gang-of-four idea can be generalized to a wide variety of games, so that even if you held the unorthodox view that income maximization alone necessarily requires the play of a subgame-perfect equilibrium, you would have to concede that

the introduction of small perturbations to the game under study can force you to change your prediction to some other Nash equilibrium of the original game. (See Fudenberg, Kreps and Levine [17].) Moreover, evolutionary modeling shows that it is very easy indeed for an adaptive process to converge on Nash equilibria that are not subgame-perfect, or which are weakly dominated. (See Samuelson [25].)

Game theorists were admittedly sold on the idea of subgame-perfect equilibrium some twenty years or so ago, but theoretical results like those mentioned above have led to the idea falling into disfavor. The modern view is that no Nash equilibria can safely be eliminated by appealing to "rationality refinements" that go beyond the assumption that both players know the other is a maximizer of utility. This theoretical rejection of backward induction is supported by a large body of experimental papers, the most convincing of which is that of Camerer *et al* [13]. Backward induction does not even work in two-stage Ultimatum Games, when one is allowed to attribute to the players the kind of inequity-averse social preferences favored by Fehr and Schmidt (Binmore *et al* [7]).

There remain economic theorists like Aumann [1] who defend backward induction as rational play in very idealized circumstances, but nobody, including Aumann and his followers, believes in using subgame-perfection for predictive purposes in laboratory experiments.

The seemingly interminable multiplication of cases in which backward induction fails to predict human behavior in the laboratory is therefore pointless. The claim that income maximization entails backward induction is not so much a straw man as a dead horse. I do not know of any economist using income maximization for predictive purposes in their experiments who feels at all threatened when shown yet another example in which backward induction fails. They all know that you need a lot more than straight utility maximization to justify backward induction.

6 Ultimatum Game

All these issues come to a head in any discussion of the Ultimatum Game, which is the game on which the authors of *"Economic Man" in Cross-Cultural Perspective* concentrate. I hope to explain here why it would be hard to find a game to study in the laboratory less well suited to refuting orthodox economic theory if that were your aim.

In the Ultimatum Game, a sum of money can be divided between Alice and Bob if they can agree on a division. The rules are that Alice proposes a division and that Bob is then restricted to accepting or refusing. The game was originally proposed by Reinhard Selten—the inventor of subgame-perfect equilibrium—to his student Werner Güth as an example in which subgame-perfection would be unlikely to work in the laboratory. Güth [18] and his coworkers confirmed Selten's intuition, and thereby created a small experimental industry in which their results are endlessly replicated.

If the subgame-perfect equilibrium (in which Bob acquiesces when Alice demands

almost all the money) were the only Nash equilibrium of the game, then the fact that Alice's modal offer in the laboratory is a fifty:fifty split would be a serious challenge to game theory, since this conclusion is indeed robust when the amount of money is made large or repeated play (against a new opponent each time) is allowed.

However, the Ultimatum Game actually has many Nash equilibria. In fact, any split of the money whatsoever is a Nash equilibrium outcome on the income maximizing hypothesis. Not only does the Ultimatum Game have many Nash equilibria, but computer simulations show that plausible models of adaptive learning can easily converge on one of the infinite number of Nash equilibria that are not subgame-perfect (Binmore, Gale and Samuelson [6]).

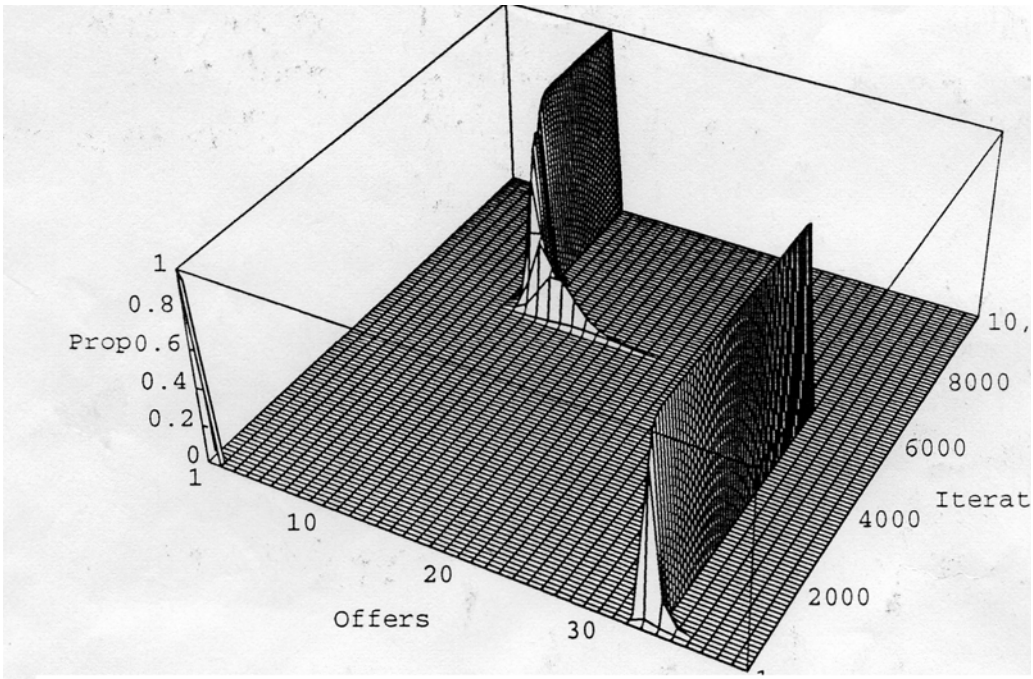
The same computer simulations show that one must expect any convergence that takes place to be very slow. (See also Roth and Erev [23].) Figure 1 shows one of the very large number of computer simulations reported by Binmore $et\ al$ [6].

The original sum of money is \$40 and the simulation begins with Alice offering Bob about \$33, leaving \$7 for herself. One has to imagine that the operant social norm in the society from which Alice and Bob are drawn selects this Nash equilibrium outcome from all those available when ultimatum situations arise in their repeated game of life. However, this split (like any other split) is also a Nash equilibrium outcome in the one-shot Ultimatum Game.

The figure shows our (perturbed replicator) dynamic leading the system away from the vicinity of this $(7, 33)$ equilibrium. The system eventually ends up at a $(30, 10)$ equilibrium. This final equilibrium is not subgame-perfect (where the split would be $(40, 0)$), but this fact is not the point of drawing attention to the simulation. What is important here is that it takes some 6,000 periods before our simulated adaptive process moves the system any significant distance from the vicinity of the original $(7, 33)$ equilibrium. This enormous number of periods has to be compared with the 10 or so commonly considered "ample" for adaptive learning to take place in the laboratory.

More generally, if a society's social norms lead inexperienced players to start playing close to a Nash equilibrium of a one-shot laboratory game, then, if there is any movement away from the original Nash equililibrium at all due to adaptive learning, we must expect such movement to be very slow at the outset. As a consequence, the much replicated results in the Ultimatum Game represent no threat to the income-maximizing hypothesis.

To threaten the income-maximizing hypothesis, one needs to use a game that does not share the pathologies of the Ultimatum Game described above. Such a game would not confuse the issue by having a huge number of Nash equilibria. Nor would simple models of adaptive learning converge only with glacial slowness. The Prisoners' Dilemma meets both criteria without difficulty—but we have seen that 90% of subjects eventually end up acting as though they were maximizing their income in the one-shot Prisoners' Dilemma.


1
0.8
Prop0.6
0.4
0.2
0
1
10
20
Offers
30
10,
8000
6000
Iterat
4000
2000

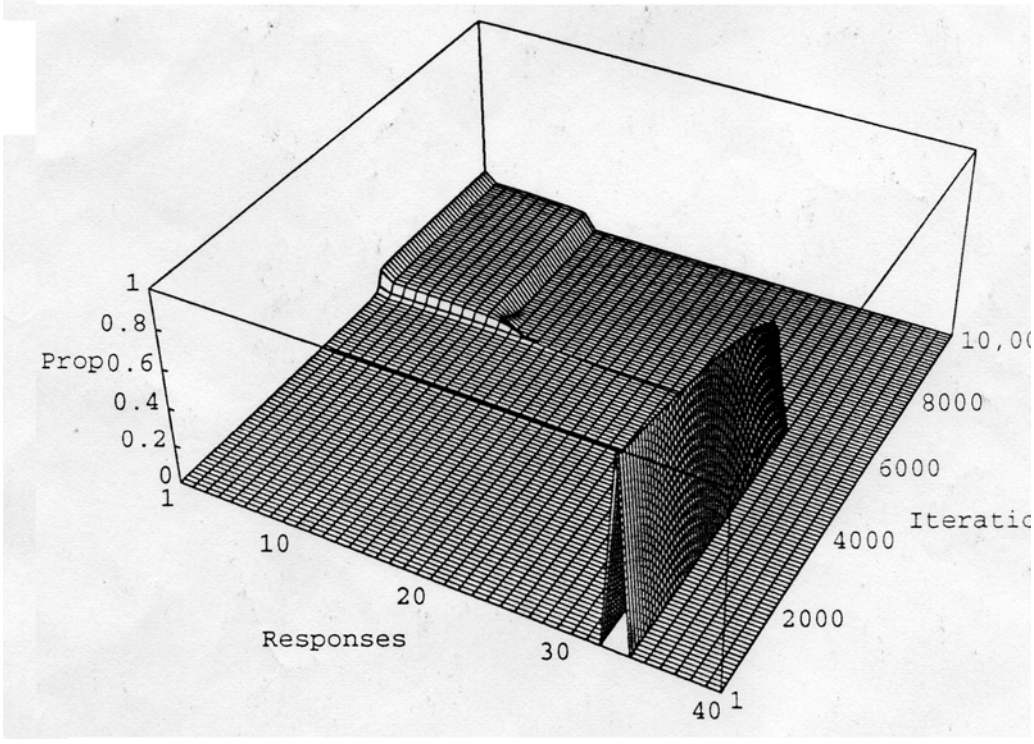

1
0.8
Prop0.6
0.4
0.2
0
1
10
20
Responses
30
40
1
10,0
8000
6000
Iteratic
4000
2000

Figure 1

7 Conclusion

Economic theory is a popular target for those who would like to live in a fairer world, but we need not believe the claims made by irresponsible neo-conservative economists when they claim that equity is incompatible with efficiency. We certainly do not need to attack the principles of orthodox economic theory in order to show that they are wrong. It is even less necessary to follow the authors of the rhetoric in *"Economic Man" in Cross-Cultural Perspective* and *Foundations of Human Sociality* by misrepresenting what these principles are.

It is an enormous pity that the gathering of the kind of quantitative anthropological data reported in these works should have been left until many traditional societies are on their last legs, and others have vanished altogether. It would be an even greater pity if the small amount of data available were to be discredited by the way it is presented.

References

[1] R. Aumann. Backward induction and common knowledge of rationality. *Games and Economic Behavior*, 8:6–19, 1995.

[2] R. Axelrod. *The Evolution of Cooperation*. Basic Books, New York, 1984.

[3] K. Binmore. *Playing Fair: Game Theory and the Social Contract I*. MIT Press, Cambridge, MA, 1994.

[4] K. Binmore. *Just Playing: Game Theory and the Social Contract II*. MIT Press, Cambridge, MA, 1998.

[5] K. Binmore. *Natural Justice*. Oxford University Press, New York, 2005.

[6] K. Binmore, J. Gale, and L. Samuelson. Learning to be imperfect: The Ultimatum Game. *Games and Economic Behavior*, 8:56–90, 1995.

[7] K. Binmore, J. McCarthy, L. Samuelson, and A. Shaked. A backward induction experiment. *Journal of Economic Theory*, 104:48–88, 2002.

[8] K. Binmore, J. Swierzsbinski, S. Hsu, and C. Proulx. Focal points and bargaining. *International Journal of Game Theory*, 22:381–409, 1993.

[9] K. Binmore, J. Swierzsbinski, and C. Proulx. Does minimax work? an experimental study. *Economic Journal*, 111:445–465, 2001.

[10] R. Boyd and P. Richerson. *Culture and the Evolutionary Process*. University of Chicago Press, Chicago, 1985.

[11] R. Boyd and P. Richerson. Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethnology and Sociobiology*, 13:171–195, 1992.

[12] C. Camerer. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press, Princeton, NJ, 2003.

[13] C. Camerer, E. Johnson, T. Rymon, and S. Sen. Cognition and framing in sequential bargaining for gains and losses. In A. Kirman K. Binmore and P. Tani, editors, *Frontiers of Game Theory*. MIT Press, Cambridge, MA, 1994.

[14] E. Fehr and S. Gachter. Cooperation and punishment in public goods experiments. *American Economic Review*, 90:980–994, 2000.

[15] E. Fehr and K. Schmidt. A theory of fairness, competition and cooperation. *Quarterly Journal of Economics*, 114:817–868, 1999.

[16] E. Fehr and K. Schmidt. The rhetoric of inequity aversion—a reply. See http://www.wiwi.uni-bonn.de/shaked/rhetoric/, 2005.

[17] D. Fudenberg, D. Kreps, and D. Levine. On the robustness of equilibrium refinements. *Journal of Economic Theory*, 44:354–380, 1988.

[18] W. Guth, R. Schmittberger, and B. Schwarze. An experimental analysis of ultimatum bargaining. *Journal of Behavior and Organization*, 3:367–388, 1982.

[19] J. Henrich *et al*. *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*. Oxford University Press, New York, 2004.

[20] J. Henrich *et al*. "Economic man" in cross-cultural perspective. to appear in *Behavioral and Brain Sciences*, 2005.

[21] D. Kreps, P. Milgrom, J. Roberts, and R. Wilson. Rational cooperation in the finitely repeated Prisoners' Dilemma. *Journal of Economic Theory*, 27:245–252, 1982.

[22] J. Ledyard. Public goods: A survey of experimental research. In J. Kagel and A. Roth, editors, *Handbook of Experimental Game Theory*. Princeton University Press, Princeton, 1995.

[23] A. Roth and I. Erev. Learning in extensive-form games: Experimental data and simple dynamic models in the medium term. *Games and Economic Behavior*, 8:164–212, 1995.

[24] D. Sally. Conversation and cooperation in social dilemmas: A meta-analysis of experiments from 1958 to 1992. *Rationality and Society*, 7:58–92, 1995.

[25] L. Samuelson. Does evolution eliminate dominated strategies? In A. Kirman K. Binmore and P. Tani, editors, *The Frontiers of Game Theory*. MIT Press, Cambridge, MA, 1994.

[26] L. Samuelson. *Evolutionary Games and Equilibrium Selection*. MIT Press, Cambridge, MA, 1997.

[27] L. Samuelson. Foundations of human sociality: A review essay. To appear in *Journal of Economic Literature.*, 2005.

[28] R. Selten and R. Stocker. End behavior in finite sequences of prisoners' dilemma supergames: A learning theory approach. *Journal of Economic Behavior and Organization*, 7:47–70, 1986.

[29] A. Shaked. The rhetoric of inequity aversion. See http://www.wiwi.uni-bonn.de/shaked/rhetoric/, 2005.

[30] R. Trivers. The evolution of reciprocal altruism. *Quarterly Review of Biology*, 46:35–56, 1971.

[31] P. Young. The evolution of conventions. *Econometrica*, 61:57–84, 1993.