Abstract

Experimental economists and other social scientists have discovered an important form of human behavior that has been inadequately analyzed by behavioral scientists. In public goods, ultimatum, and other games where players gain from cooperative behavior, agents have a predisposition to cooperate and to undertake costly punishment of defectors, even when this behavior cannot be justified in terms of traditional game-theoretic equilibrium and learning concepts. We call this 'reciprocal fairness.'

Our research has four goals. First, can the experiments on reciprocal fairness be replicated with diverse subject pools and various strategic settings? Second, how might such behavior have evolved, given that it is formally altruistic, and hence 'unfit' except under stringent circumstances? Third, how does the existence of reciprocal fairness influence our analysis of social policy in such areas as taxation, charity, redistributive expenditure, and criminal sentencing? Fourth, how to what extent does cultural variation induce differences in the strength of reciprocal behavior and conditions under which agents exhibit reciprocal fairness?

1 Explaining Cooperative Behavior
As Bernard de Mandeville and Adam Smith long ago observed, the proper institutional framework can induce self-interested agents to serve the interest of others. The most sophisticated modern statement of this principle is probably the Fundamental Theorem of Welfare Economics, based on the Walras-Arrow-Debreu general equilibrium model.

This `neo classical' explanation of cooperation presumes that all economic transactions are fully contractible, and all contracts are costlessly enforceable. However, cooperation in modern market economies is not limited to situations of complete and costlessly enforced contracts. When the assumptions of the general equilibrium model are appropriately weakened a different set of analytical tools are needed to explain why self-interested agents cooperate.

One such tool is the repeated game, which implies the famous `Folk Theorem,' implying that when discount rates are sufficiently low, Pareto-optimal cooperation can be sustained if each agent responds to a non-cooperative action by refusing to cooperate for the remainder of the game. Under suitable conditions such threats can be shown to be credible, in the sense that it will indeed be in the self-interest of each agent to carry out the threat of retaliatory defection should the occasion to do so arise, given that every other individual does so as well.

While few instances of social cooperation actually use universal defection to sustain cooperation, other plausible sanctions can serve the job as well.

The principal-agent model and its many variations have also been used widely to explain cooperation in the crm, and have been extended to provide an analytical basis for a theory of economic institutions based on treating individuals as rational and self-interested.

However, important forms of cooperative behavior are commonly observed, for instance, Debreu (1959), Arrow and Debreu (1954), Arrow and Hahn (1971), or Mas-Colell (1985).


One is that people are "irrationally" prosocial: in laboratory settings subjects consistently appear to contribute more than the rational actor model predicts in public goods games (Marwell and Ames 1979, Schneider and Pommerehne 1981, Dawes, de Kragt and Orbell 1988, Isaac and Walker 1988b, Isaac and Walker 1988a). For a review of the extensive literature on this subject, see Ledyard (1995). We may summarize this research as follows. Only a fraction of subjects consistently defect, contributing nothing to the public account. Rather, people begin by making contributions that average about midway between the perfectly cooperative and the perfectly noncooperative levels, although cooperation deteriorates if the game is repeated numerous times.

There have been some attempts to explain prosocial public good contribution on the basis of the rational actor model, most notably Joel Guttman (1986, 1987). The models we have seen, however, require implausible commitment assumptions, and do not explain the laboratory evidence on the public goods game. Others have attempted to reconcile this behavior with "rationality" by noting that in repeated public goods games, cooperation decreases over time, eventually approximating the unique subgame perfect equilibrium. This suggests that participants do not understand the game at first, but progressively learn the superiority of the free-riding strategy.

There is persuasive evidence that this interpretation is incorrect, however. For instance, Andreoni and Fehr and Tyran (1996, 1996a, 1996b, 1996c) found that when such games are repeated with the same subjects, the initial levels of cooperation are restored, but once again cooperation decays as the game progresses.

These authors suggest that the decay of cooperation in repeated public goods games occurs because public-spirited contributors retaliate against free-riders in the only way available to them in the game: by defecting themselves.


Some have explained this anomaly simply as error on the part of responders (for an overview, see Davis and Holt 1993 and Fehr and Tyran 1996).
2 Homo reciprocans

reciprocal fairness

\footnote{One of us (Herbert Gintis) has used artificial life simulations of the ultimatum game indicating that sufficiently high mutation rates and sufficiently low rates of migration among groups (or a sufficiently small size of the whole population) can reproduce the empirical finding of the ultimatum game. This result, however, is more plausibly interpreted as genetically-evolved vindictiveness rather than ‘noise’ or ‘error’ on the part of responders, since normally a considerable fraction of the population exhibits retaliatory behavior.}

\footnote{In the work of Ostrom et al. (1992) the same group of subjects interacted for roughly 25 periods, and subjects could develop an individual reputation for punishing defectors. Their experimental design therefore permits an interpretation of costly retaliation in terms of strategically rational behavior: retaliation may increase cooperation in future periods. In Fehr and Gächter (1996), group composition is changed in every period and individual reputation formation is ruled out by the design. Therefore, costly retaliation does not}
respond to defection with defection. We dub such agents Homo reciprocans, to highlight the contrast of this behavior with that of the more traditional Homo economicus. Homo reciprocans is thus neither the sel/n0dess altruist of utopian theory, nor the sel/n0csh hedonist of neo-classical economics. Rather, he is a conditional cooperator whose penchant for reciprocity can be elicited under the proper circumstances.

The novel element in this research is not the recognition that reciprocal behavior exists. The importance of altruism within families has been stressed by William Hamilton in his seminal work on `inclusive fitness', and Robert Trivers has shown that no concept of rationality is need to predict that even unrelated individuals can gain from reciprocal behavior in repeated interactions. The robustness of reciprocal behavior appears in computer simulations as well, as in the work of Hamilton and Robert Axelrod, reviewed in Axelrod and Dion. Artificial life simulations of repeated prisoner’s dilemma games using genetic algorithms also show the robustness of strategies that are `nice’—never defect, `punishing’ always punish defection and `forgiving’: return to cooperation after a short period of punishing, if the other player is cooperating.

Rather, the novel element is the fact that retribution is `altruistic’ in the sense that retaliatory behavior benefits the group by fostering sustained cooperation, but at a strictly positive cost to the individuals who bear the trait.

3 Topic I: Experimental Research on Sustaining Cooperation Via Reciprocal Fairness

Homo reciprocans

Homo economicus  Homo reciprocans

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.

Levine (1996) analyzes a utility function that combines cooperation and retribution.

Non-reciprocating strategies also emerge from artificial life simulations. On the fitness of non-reciprocating strategies, see Nowak and Sigmund (1993).
Research Initiative on Reciprocal Fairness June /July, 1997

Resources are wasted for retaliation. This result raises, however, the question to what extent the uncontrollable effects of face to face communication alone can account for the increase in cooperation and welfare. One important task of the planned experiments is, therefore, to determine which coordination devices facilitate the achievement of welfare gains through the threat of costly retaliation.

Reciprocal fairness is not only characterized by conditional retaliation but also by conditional kindness. The Ostrom et al. and Fehr-Gnechter experiments allow, however, only for retaliation. An interesting question is whether the combination of the opportunity to exchange "gifts" with the opportunity to retaliate leads to improved outcomes in terms of rates of cooperation and the welfare of group members. Previous research indicates that in bilateral relations, positive reciprocity is generally associated with welfare gains. Whether this also holds in the context of n-person social dilemmas remains an open question.

In the Ostrom et al. and the Fehr-Gnechter experiments groups were relatively small and all group members could monitor and punish the cooperation behavior of all the other group members. This design assumes implicitly a rather dense social structure: every agent can monitor and punish every other agent in the group. In fact, people often interact with a relatively small number of neighbors, so the structure of social interaction is not as dense as assumed in the laboratory experiments. We will investigate how variations in the density of social interaction affect cooperation rates. Such experiments can help to detect those social structures that enhance cooperation and those that inhibit cooperation.

Further questions concern the effects of heterogeneity in endowments and preferences on cooperation. Does equality in endowments enhance cooperation? Does heterogeneity of preferences inhibit cooperation? To what extent do those with a strong preference for the public good force those with a weak preference to contribute? How are cooperation rates and welfare affected if heterogeneity in preferences is combined with mobility across groups? Virtually all of these questions are not yet rigorously examined, although they are of fundamental importance if we want to better understand which social structures and institutions will foster cooperative outcomes.

Another relevant issue is how social structure affects cooperation. We hypothesize that the more dense the social interaction among agents, the more effective is costly punishment in maintaining cooperation. Consider, for instance, a public goods game with 20 subjects. The return of one token on the public account is 0.04. So if each subject invests the whole...
Research Initiative on Reciprocal Fairness June /7/, /1/9/7

Endowment, subjects can double their income relative to a situation where each puts the whole endowment in the private account. Treatment A: Each subject can punish all the other /2/4 players in the game. Treatment B: Each subject can only monitor and punish a subset of the other players; i.e., there exist "local interactions." The number of agents who can be monitored and punished by each person in the group is thus a proxy for social density or social capital. By varying the social density we can study its impact on cooperation. This may be a particularly important contribution to the current policy analysis of the role of "social capital" in inducing cooperative behavior in communities. This same research design permits the study of the "in/spatial" diffusion of cooperation. Suppose, for example, that subjects are spatially ordered on a rectangular grid. Then those in the middle have the largest number of social contacts while those at the corners have a low number of contacts. In the corners it may therefore be easier to free-ride without being punished. Hence, in the corners free-riding may survive while in the center cooperation is maintained.

Finally, we intend to support the extension of experiments concerning reciprocal fairness to social interactions relevant to real-world social policy issues, including charitable contributions, taxation, redistribution in the "welfare state," the treatment of social inequality, and criminal sentencing. Such experiments involve adding contextual and interpretive elements to the bare structure of game payoffs. While such additions can lead to a situation where experimental outcomes are subject to multiple interpretations, we believe carefully limiting and controlling the contextual material can avoid this problem.

4 Topic II: Explaining Reciprocal Fairness

culture of honor
Evidence that reciprocal fairness is a ubiquitous behavioral pattern, should the research of topic I support this hypothesis, then poses a puzzle. How could costly punishment and other forms of reciprocal fairness evolve? The problem is not new; Hume rhetorically asked:

Who sees not that vengeance, from the force alone of passion, may be so eagerly pursued as to make us knowingly neglect every consideration of ease, interest, or safety?

Formally, the puzzle is this. Costly punishment, and the various other forms of reciprocal fairness need not be, and generally are not in the interest of the actor at the time the action is to be taken. If actors have outcome-based preferences that are entirely self-regarding, the threat to retaliate is not generally credible, and an equilibrium in which behaviors characterized by reciprocal fairness may not be subgame perfect. An obvious resolution of the puzzle is to posit that people take pleasure in indicting harm on those who have done harm to them or to others or are driven to do so by a sense of obligation, honor, or dignify. But this simply displaces the puzzle. How could preferences or compulsions of this type have evolved?

We will attempt to provide an answer using models of genetic evolution, cultural evolution, and gene-culture co-evolution. Thus we will support the development to evolutionary models of preference formation that determine conditions under which reciprocal fairness might emerge. Such research involves relatively uncharted territory, although there has been some research into the evolutionary emergence of the rate of time preference, degree of risk aversion, and altruistic behavior, using group selection arguments (Hansson and Stuart, 1990, Rogers, 1994, Mailath, Samuelson, and Shaked, 1995, Robson, 1995), and kin selection (Bergstrom and Stark, 1993, Bergstrom, 1995).

An obvious candidate for a mechanism explaining the emergence of reciprocal preferences is group selection operation on either cultural or genetic transmission mechanisms. Why might some genetic basis be contemplated for so complex a social behavior as reciprocal fairness? The answer is that the most notable attempt to explain retaliation is doubtless Sethi and Somanathan (1996), who use neither group selection, nor local interactions, nor other forms of heterogeneity to prove that under the appropriate conditions costly retaliation against defectors can sustain a cooperative equilibrium in a common pool resource game. However their result depends on the absence of mutation or what we consider to be implausible patterns of mutation—in particular, that cooperators mutate into retaliators at a sufficiently high rate to squelch the emergence of defectors. More important, their model implies a very low level of defection and retaliation, whereas we believe human societies exhibit high levels of defection and very high levels of reciprocal fairness. 

\footnote{The most notable attempt to explain retaliation is doubtless Sethi and Somanathan (1996), who use neither group selection, nor local interactions, nor other forms of heterogeneity to prove that under the appropriate conditions costly retaliation against defectors can sustain a cooperative equilibrium in a common pool resource game. However their result depends on the absence of mutation or what we consider to be implausible patterns of mutation—in particular, that cooperators mutate into retaliators at a sufficiently high rate to squelch the emergence of defectors. More important, their model implies a very low level of defection and retaliation, whereas we believe human societies exhibit high levels of defection and very high levels of reciprocal fairness.
formal institutions involved in the transmission of culture in Western societies do not "teach" reciprocal fairness. The norm of "returning kindness with kindness" is widely shared and promulgated, but the norm of "returning evil for evil" is held to betray a low level of moral reasoning and behavior. Indeed, few forms of retaliatory behavior are considered praiseworthy, although some are widely understood and tolerated. For instance, we sympathize with the emotions of the victim of a crime to increase the severity of punishment imposed upon the perpetrator of the crime. By contrast, of course, victims of crimes are praised for their pro-social efforts to reduce the general level of a type of criminal activity. This however is not a case of reciprocal fairness, since it could easily be understood in terms of a reputation effect: people are rewarded for being "good citizens" in opposing anti-social activity. While biologists have expressed considerable skepticism concerning the ability of group selection arguments to explain altruism in most species, there may be unusual characteristics of Homo sapiens and our close ancestors allowing biological group selection to work with greater force. Among these distinct characteristics is the superior ability of Homo sapiens to maintain group membership boundaries and practice exclusion thus reducing the level of intergroup mobility and enhancing the force of group selection. Similarly, while recent empirical work on cultural group selection (Soltis, Boyd and Richerson 1995) suggests that the process may work very slowly, we are not persuaded that this counts as an argument against the force of cultural group selection for traits of the kind we are studying. Further, as the proposed research of Robert Boyd suggests, the assertion that cultural group selection is moving may be a model-specific result stemming from particular assumptions concerning intergroup migration. To address these possibilities, we will extend the group selection work of Boyd and his coauthors, combining it with other research (Bowles and Gintis 1997, Bowles and Gintis 1998b) concerning endogenous group formation, boundary maintenance, and the evolution of cooperation within groups. We will also explore other mechanisms, including variants of nonrandom pairing of agents other than group selection (e.g., local interaction see Mailath et al. 1995 and Herbert Simon's "pleiotropic" model, whereby mathematical biologists have shown that the conditions favoring group selection of altruistic behavior are extremely restrictive, and the behavior of most life forms can be explained without recourse to dynamics based upon altruism or group selection (Williams 1966, Dawkins 1989, Maynard Smith 1976). Nevertheless eusocial nonhuman species have emerged and do very well, and group selection is part of the account of their existence. See Alcock (1993), Ch. 16, for a recent review of the literature. For the case of Homo sapiens see Caporael (1987), Simon (1993) and the discussion in Wilson and Sober (1994).
5 Topic III: Cross-Cultural Research in Reciprocal Fairness
Research Initiative on Reciprocal Fairness
June 7, 1997

them in the theory and methodology of the ultimatum game and other related experimental games, devise a set of experimental variations to test specific hypotheses, and send them to their field sites across the globe with a common research design intended to acquire cross-cultural, comparative data on ultimatum game performance. When all the researchers have completed their investigation, we will reconvene, share our data, and discuss our analyses. The results and conclusions of this conference, which will be published in an integrated form in an edited volume, should allow us to begin differentiating and delineating the aspects or components of ultimatum game behavior that result from innate pan-human cognitive processes from those aspects that rely on culturally evolved behavioral norms or rules.

In the ultimatum game, two players are allotted a sum of money. The first player, often called the proposer, offers a portion of the total sum to a second person, often called the `responder.' The responder, then, has the opportunity to accept or reject the proposer's offer. If the responder accepts, she receives the amount of the offer and the proposer receives the remainder. If the responder rejects, then nobody gets anything; both the responder and proposer receive zero.

The ultimatum game first sparked the interest of economists because its highly replicable results radically departed from the standard behavior of self-interested income maximizers. Game theory unambiguously predicts that proposers will offer the smallest non-zero amount possible, and responders will always accept. Instead, offers made by proposers typically average between 30 and 40 percent of total, with the modal offer at 50 percent. Responders usually accept the average offers, and will often reject offers lower than 20 percent (Camerer and Thaler 1995).

Experimental economists have systematically studied the influence of various factors on the game's results, including stake size, degree of anonymity, context and `culture.' Lisa Cameron's (1995) analysis of data from Indonesia, where she was able to provide sums equivalent to approximately three month salary for test subjects, strongly rejects the hypothesis that higher stakes move individuals closer to game-theoretic behavior. In fact, her data suggest that proposers generally move away from game-theoretical predictions and toward a 50%-50 split; responders, consequently, accept these proportionately higher offers more frequently. Similarly, Hoebman, McCabe and Smith (1994) tested the effect of raising the stakes from $25 to $250 dollars, and found they could not reject the hypothesis that the offers are identical with $250 stakes and with $2500 stakes.

Researchers have also suggested that the experimenter's knowledge of the proposer's behavior may contribute to non-equilibrium `non-rational'...
5.1 The Effect of Environmental Differences on Behaviors
Research Initiative on Reciprocal Fairness June/July 1997

And, proposers in Pennsylvania who provided a mode of , and, Israeli responders were, on average, willing to accept somewhat lower numbers than students at the University of Pittsburgh. Even Cameron’s extensive data from students and faculty working at Gadjah Mada University in Yogyakarta, Indonesia, perhaps the best cross-cultural test of the ultimatum game, revealed no significant differences in comparison to data from Roth et al. or Hoffman et al. The mean proposer demands, for example, from Indonesia and the U.S. were 0.573 and 0.562, respectively.

In total, many factors generate small deviations in ultimatum game performance, but no alterations in experimental variables have produced substantial deviations from the usual results except for fundamental changes like removing punishment possibilities from the game. After his review of these studies in the Handbook of Experimental Economics, Alvin Roth concludes the following:

Thus we see here a series of experiments whose results seem to be that even initially very skeptical investigators are becoming persuaded that the experimental results observed in ultimatum games are not easily displaced artifacts of the experimental methods, but rather represent very robust phenomenon.

Interestingly, in his cross-cultural analysis of the small differences between American and Israeli performance, Roth suggests that these apparent differences indicate, not a difference in aggressiveness or toughness, but rather a difference in what is perceived as fair, or what is expected under the circumstances. Some economists might suggest that these minor differences in what is ‘expected’ or ‘fair’ between cultural groups result from a sort of window-dressing effect that cultural beliefs can have on more fundamental, innate economic reasoning processes. Some new research supports Roth’s conclusions, and indicates that culture may be an important variable in understanding game performance.

Recently, an anthropologist Joel Henrich, working with Robert Boyd, conducted an experiment among the Machiguenga, an Arawak-speaking indigenous group inhabiting the southeastern Peruvian Amazon, which suggests that cultural differences can have a substantial effect on behavior in the ultimatum game. Traditionally, the Machiguenga live in mobile, single family units or small family hamlets and subsist on a combination of swidden (sometimes called slash and burn) agriculture, hunting, gathering and fishing. Within the last 30 years, missionaries, government-sponsored bilingual schools, and markets have sedentized and centralized the Machiguenga in a gradual process of increasing market integration. Currently,
most Machiguenga live in small semi-permanent agricultural communities between 250-350 people, grow some cash crops, and subsist primarily on manioc, plantains and some.

In the summer of 1996, during Henrich’s third visit to the Machiguenga community of Camisea, he performed a moded version of the ultimatum game experiment. First, he gathered twelve men together between the ages of 18 and 30 under the auspices of “playing a fun game for money.” He explained the game to the group in Spanish using a set script written with simple terminology like “first person” to reference the proposer and “second person” for the responder. After this he had a bilingual school teacher re-explain the game in the Machiguenga language translating from his script, and display the money that he would be using to make payments. After this, each participant entered Henrich’s house the guest hut individually, he and the teacher explained the game a third time, and Henrich asked a number of practice hypothetical questions intended to test the participants’ comprehension of the game. They re-explained parts of the game as necessary. After the individual consistently answered at least two hypothetical questions correctly, he would submit the actual question with a pile of soles Peruvian money in view. The following day, after having successfully gotten 12 responses and paid out some money, he began seeking randomly selected individuals to play the game. Most people had already heard of the game and were eager to play. He privately explained the game to each individual usually in their house and ran through the same testing procedure as the previous day.

After three days of doing this he accumulated 32 responses. During this process several people were rejected because they, after 30 minutes of explanation, could not understand the game at least they could not answer the hypothetical questions. While typical U.S. results produce a mean of 40%, a mode of 50%, and few numbers below 20%, the Machiguenga proposers proposed a mean number of 27./5%, with a mode of 25%, and many numbers of 15%. Similarly, Machiguenga responders, with one exception, always accepted; many numbers of 15% were accepted, whereas, Americans frequently reject numbers below 20%. These results seem to be very different from what has been observed elsewhere.

We hypothesize that the Machiguenga behave differently from subjects in other experiments because they are culturally different from those subjects; that is, because they have socially learned different values, beliefs and behaviors from preceding generations. We believe that cultural effects were not detected in previous experiments because the subjects in prior experiments were in fact culturally very similar; all were urban university
5.2 Research Method in a Small Scale Society
would not serve the same purpose as the proposed control experiment. The proposed experiment is meant to control for as much as possible, including experimenters and the specifics of a cross-cultural protocol, not the effect of stakes. Moreover, given the cost of Amazonian research, redoing the Machiguenga study would also be much more expensive.

In comparison to most previous ultimatum game procedures, the Machiguenga experiment used relatively high stakes. The base sum was 20 soles (~28/29), which represents about 2.2 days pay for a Machiguenga male doing wage labor. We propose ordering graduate students approximately 2510, or roughly 2.2 days pay readers, for example, make about 259.50. This sets the stakes high in comparison to the usual amount, and approximately equal to the Machiguenga experiment. If the Machiguenga were only responding to high stakes, then UCLA graduate students should perform similarly.

The Machiguenga seemed confident of anonymity among participants; in fact, they did not seem to care about anonymity at all, but each participant knew she or he was playing against someone else in a small community of approximately 70 adults, so it was a guarantee they knew the other player, they just did not know specifically who it was. Again, by using only UCLA anthropology graduate students, which number approximately 80, and by making this known to all participants, we hope to create a comparable social situation. Participants will be assured of anonymity, but they will know they are playing against someone else they know.

The 32 student participants will be randomly selected from a stratified list of all anthropology graduate students to match gender and age differences recorded in the Machiguenga sample. That is, the total list of anthropology graduate students will be subdivided by gender and age. The appropriate number of participants will be randomly selected from each age/gender subdivision. This should mitigate any argument that differences in performance between Machiguenga and UCLA students arise from variations in the gender/age composition of the samples.

All other conditions of the Machiguenga experiment will be approximated as closely as possible. We will use the identical game description, terminology and explanatory examples except in English, rather than Spanish or Machiguenga. In designing the Machiguenga experiment Henrich intentionally avoided any complex or suggestive terminology like `buyer' and `seller' or `proposer' and `responder,' and instead opted for labels like `1st person' and `2nd person.' Because he initially avoided words imbued with implicit cultural assumptions, translations should proceed smoothly and with little distortion. As he did with the Machiguenga, he plans to...
5.3 Expanded Cross-Cultural Research
Medium of exchange.
The value of money, in terms of what it can buy, the social rules for meat distribution may be quite different from the rules for cash distribution.

The sphere of exchange may influence game performance. Although the game is usually played anonymously, people may have implicit assumptions about the pool of potential participants with which they are re-playing. Most experiments are performed at universities, so student participants may accurately assume they are paired with another student, and perhaps a classmate. In the Machiguenga experiment, individuals were told that they played with someone else in their community of 300 people. How would it affect the results if we controlled for players' assumptions about the sphere of exchange? Do cultural differences affect how people react to different spheres? Many villages in Amazonian, Africa and New Guinea, for example, are subdivided into clans or moieties. What if we changed the sphere from the village (say about 500 people) to the clan (150 people) or moiety (250 people)? Does the sphere affect the tendency to make equal divisions or the desire to punish 'unfair' offers? How does ethnicity affect the equation: What if we told Mongolian pastoralists they were playing against neighboring but ethnically distinct Kazakhs? Would they play differently against other Mongols, then against anonymous Kazakhs?

Situational characteristics.

Sphere of exchange.

Situational characteristics.
Economic and ecological correlates.
6 Topic IV: The Implications for Social Policy

Injustice: The Social Bases of Obedience

and Revolt
by no means excludes hierarchy and authority, where exceptional qualities and defects can be the source of enormous admiration and awe. At the same time, it is one where services and favors, trust and abjection, in the course of mutual exchanges, are ideally expected to condition some rough balancing out.

Morris termed the general ground plan he uncovered the concept of reciprocity, or better, mutual obligation, a term that does not imply equality of burdens or obligations.

One standard explanation for the fact that individuals vote for redistributive expenditure and voluntarily contribute to private charities is that the welfare of others is an argument in the utility function. However, it is well known that predictions based on this concept of altruism do not describe charitable behavior well and this form of altruism has perverse effects. Another standard explanation is insurance: people vote for redistributive expenditures that they might not currently receive, but could receive under some likely future states of nature. However, much of the support for redistributive expenditure for the poor is difficult to justify in this manner.

Can a theory of reciprocity contribute to an explanation of charitable giving and support for redistributive expenditure? Wein tend to commission two or three papers to address this issue. Potential researchers in this area are Robert Sugden and Martin Gilens, who have written extensively on the problem, and Christopher Jentsch, who uses the concept in his analysis of the "revolt against the welfare state," but does not employ formal models of reciprocal fairness.

We expect that reciprocity may be able to explain both the rise of the welfare state after World War II, and the more recent tide of opposition to redistributive policies in recent years. Specifically, in light of the experimental regularities outlined above, we suspect the following to be true as well: redistributive policies that reward people independent of whether and how much they contribute to society are considered unfair and are not supported, even if the intended recipients are otherwise worthy of support.

Samuel Bowles and Herbert Gintis have argued for a reciprocity theory of giving (Bowles and Gintis 1998c), but have not developed formal models of the phenomenon.

---

12 Samuel Bowles and Herbert Gintis have argued for a reciprocity theory of giving (Bowles and Gintis 1998c), but have not developed formal models of the phenomenon.
and even if the incidence of defection in the target population is not particularly high. This would explain the opposition to many welfare measures for the poor, particularly since such measures have been to some extent opportunistically exploited, and are thought to have facilitated various social pathologies. At the same time it explains the continuing support for social security and medicare in the United States, since the public perception is that the recipients are "deserving." Our artificial life simulations suggest that the moral bases of social policy may follow a dynamic cycle. In modeling the repeated Prisoner’s Dilemma, we have found that in periods in which levels of defection are high, reciprocal strategies, tit-for-tat, spontaneously emerge as individually successful strategy that leads to a very high level of cooperation. In this highly cooperative state, however, the attractive features of tit-for-tat disappear, and there is a movement towards unconditional cooperation that in turn invites high levels of defection and non-cooperative behavior. This state of affairs is then conducive to the return of reciprocating strategies, thus completing the dynamic cycle.

The cross-cultural research we will undertake also has important policy implications. Distinguishing the effects of culture from pan-human cognitive characteristics has important consequences for economic theory. The rational actor model carries with it the implicit assumption that people everywhere reason the same way, and the results of economic experiments are sensibly interpreted as revealing facts about these reasoning processes. However, if human economic behavior is affected by the cultural milieu, then theories of human economic reasoning that neglect of the cultural environment are incomplete. Thus, such experiments suggest that economic theory should be extended to consider the reciprocal effects of economic institutions and culturally transmitted beliefs.

Application-oriented scientists seeking to positively influence economic policy will have to model both innate human economic calculations and population-level process of cultural evolution. In general it will not be possible to predict patterns of behavior without understanding the cultural transmission of the norms, values, and rules that influence economic behavior. For example, development economists need to understand that millions of people in underdeveloped and developing nations may not respond as they anticipate because these peoples possess culturally evolved norms and values.

---

\footnote{Evidence for this view is presented in Gilens (1996). Similarly, Citrin and Green (1990) uses voting data and other statistical data to argue that only to a very small degree does self-interest explain to the political preferences of Americans in many policy areas, including cash grants based on means-tested criteria.}
Research Initiative on Reciprocal Fairness June 7, 1997

rules that contrast greatly with the norms and rules possessed by western industrial peoples; the environment in which most of these theories were generated. Further, culture is an evolutionary process, and consequently the norms and behavioral rules that govern ultimatum game performance in western society may rapidly change with the rise of novel social intuitions.

Economic theories founded on culturally-transmitted behavioral rules, but lacking any mechanisms for the temporal dynamics of those rules, will lose their explanatory power as our culture evolves and our society changes.

References


Arrow, Kenneth J. and Frank Hahn, /n5cGeneral Competitive Analysis/ /n7b/San Francisco/: Holden-Day, /1/9/7/1/n29/.

and Gerard Debreu, /n5cExistence of an Equilibrium for a Competitive Economy/, /n7b/Econometrica/ /1/9/5/4/n29/:/2/6/5/n7b/2/9/0/.


and Douglas Dion, /n5cThe Further Evolution of Cooperation/, /n7b/Science/ /2/4/2/ /1/9/8/8/n29/:/1/3/8/5/n7b/1/3/9/0/.

and William D. Hamilton, /n5cThe Evolution of Cooperation/, /n7b/Science/ /2/1/1/ /1/9/8/1/n29/:/1/3/9/0/n7b/1/3/9/6/.

Bendor, Jonathan and Dilip Mookherjee, /n5cInstitutional Structure and the Logic of Ongoing Collective Action/, /n7b/American Political Science Review/ /8/1/,/1 /n7b/March /1/9/8/7/n29/:/1/2/9/n7b/1/5/4/.

Bergstrom, Theodore C., /n5cOn the Evolution of Altruistic Ethical Rules for Siblings/, /n7b/American Economic Review/ /8/5/,/1 /n7b/March /1/9/9/5/n29/:/5/8/n7b/8/1/.

and Oded Stark, /n5cHow Altruism can Prevail in an Evolutionary Environment/, /n7b/American Economic Review/ /8/3/,/2 /n7b/May /1/9/9/3/n29/:/1/4/9/n7b/1/5/5/.

Bernheim, B. Douglas and Oded Stark, /n5cAltruism Within the Family Reconsidered/: Do Nice Guys Finish Last/, /n7b/American Economic Review/ /7/8/,/5 /n7b/December /1/9/8/8/n29/:/1/0/3/4/n7b/4/5/.

References

Animal Behavior: An Evolutionary Approach

Journal of Public Economics

General Competitive Analysis

Econometrica (The Evolution of Cooperation

Science

Science

American Political Science Review

American Economic Review

American Economic Review

American Economic Review
Games and Economic Behavior

International Journal of Game Theory

The Genetics of Altruism

American Economic Review

Journal of Economic Perspectives

Economics, Values and Organizations

Evolution & Human Behavior

Recasting Egalitarianism: New Rules for Markets, States, and Communities

Ethology and Sociobiology

Quarterly Journal of Economics

Altruism, Morality, and Economic Theory

Journal of Economic Perspectives
Evolutionary Epistemology: A Multiparadigm Program

Research in Micropolitics

The Adapted Mind: Evolutionary Psychology and the Generation of Culture

Quarterly Journal of Economics

Experimental Economics

Acta Psychologica

The Selfish Gene, 2nd Edition

Theory of Value

Journal of Economic Behavior and Organization

Nordic Journal of Political Economy

Econometrica

Games and Economic Behavior

Econometrica
| Journal of Economic Behavior and Organization | Animal Behavior |
| economistica | Journal of Political Economy |
| – | – |
| – | – |
| Biology | American Naturalist |
| – | – |
| – | – |
| American Economic Review | Journal of Theoretical Biology |
| Journal of Political Economy | Human Ecology |
| – | – |
| – | – |
| – | – |
| – | – |
Research Initiative on Reciprocal Fairness June 7, 1997


Journal of Political Economy

American Journal of Sociology

The Theory of General Economic Equilibrium: A Differentiable Approach

Quarterly Review of Biology

Injustice: The Social Bases of Obedience and Revolt

Nature

American Political Science Review

Economic Anthropology

The Handbook of Experimental Economics

American Economic Review
Researc h Initiativ e on Recipro cal F airness June /7/, /1/9/9/7

Schneider/, F riedric h and W erner W/. P ommerehne/, /n5cF ree Riding and Col-/lectiv e Action/: An Exp erimen t in Public Micro economics/, /

Quarterly Journal of Economics

Sethi/, Ra jiv and E/. Somanathan/, /n5cThe Ev olution of So cial Norms in Com/-mon Prop ert y Resource Use/, /

American Economic Review

Shapiro/, Carl and Joseph Stiglitz/, /n5cUnemplo ymen t as a W orker Discipline Dev ice/, /

American Economic Review

Simon/, Herb ert A./, /n5cAltruism and Economics/, /

American Economic Review

Soltis/, Joseph/, Rob ert Bo yd/, and P eter Ric herson/, /n5cCan Group/-functional Beha viors Ev olv eb y Cultural Group Selection/: An Empirical T est/, /

Current Anthropology

Sugden/, Rob ert/, /n5cOn the Economics of Philan throp y /, /

Economic Journal

Stark/, Oded/, /n5cOn the Economics of Nonmark et T ransfers/: the Role of Al/-truism/, /

Economic Letters

T rivers/, R/. L/./, /n5cThe Ev olution of Recipro cal Altruism/, /

Quarterly Review of Biology

Weissing/, F ranz and Elinor Ostrom/, /n5cIrrigation Institutions and the Games I rrigators Pla y/: Rule Enforcemen t without Guards/, /" in Reinhard Selten /n28ed/./n29 Game Equilibrium Models I I/: Metho ds Morals and Mark ets /n28Berlin/: Springer/-V erlag/, /1/9/9/1/n29 pp/. /1/8/8/n7b/2/6/2/.

Williams/, G/. C/./, Adaptation and Natural Selection/: A Critique of some Cur/rent Ev olutionary Though t /n28Princeton/, NJ/: Princeton Univ ersit y Press/, /1/9/6/6/n29./

Wilson/, Da vid Sloan/, The Natural Selection of P opulations and Comm uni/ties /n28Menlo P ark/, CA/: Benjamin Cummings/, /1/9/8/0/n29./

and Elliott Sob er/, /n5cRein tro ducing Group Selection to the Human Beha v/ioral Sciences/, /

Behavior and Brain Sciences