Homo Economicus and ***Zoon Politikon*****:**
Behavioral Game Theory and Political Behavior

Samuel Bowles
Herbert Gintis*

October 26, 2004

1 Introduction

The rational choice model pioneered by economists is rapidly becoming the standard approach throughout the behavioral sciences. The model is attractive as it allows the mathematical formalization of an essential truth, namely that when people act, they are generally trying to accomplish something, and their efforts are more or less effectively oriented to this end. However, its acceptance in other disciplines coincides with an increasing recognition in economics of the limitations of the behavioral assumptions sometimes summarized by the term *Homo economicus*. While *Homo economicus* is not entailed by any of the axioms of the rational choice model, in both teaching and research three assumptions embracing this behavioral model are commonly treated as integral to the approach.

First, preferences are assumed to be *outcome-regarding*; i.e., agents care about only the quantity and quality of goods and services that they possess and consume, not about the social process through which their economic opportunities are determined. In fact, preferences are also in part *process-regarding*; agents care about how they treat and are treated by others. In evaluating states, people care how those states come to be available. In particular, people care about fairness and reciprocity. Second, preferences are assumed to be *self-regarding*: agent are assumed to care only about states experienced by themselves, not by others. In fact, however, preferences are in part *other-regarding*; agents care about the well-being of others, both positively and negatively. In particular, people reward and punish the behavior of others even at a net cost to themselves.

*Affiliations: Bowles, University of Siena and Santa Fe Institute; Gintis, Santa Fe Institute and Central European University, Budapest. We would like to thank Robert Goodin, Charles Tilly, and Elisabeth Wood for comments, as well as the John D. and Catherine T. MacArthur Foundation and the Behavioral Sciences Program of the Santa Fe Institute for financial support.

Third, preferences are assumed to be either unchanging, or to evolve under influences external to the social system under consideration. While a handy—even indispensable—assumption for many analytical tasks, the assumption of *exogenous preferences* is strongly counter-intuitive, while the *social formation of preferences*, as we will see, is strongly suggested by recent behavioral experiments.

Since Aristotle introduced the idea of *zoon politikon*, students of political behavior have recognized the importance of process-regarding, other-regarding and endogenous preferences in explaining such essential aspects of political behavior as the maintenance of social order, collective action to achieve common ends, political violence, and even the simple act of voting. Recent experimental research has confirmed the existence of process-regarding and other-regarding preferences. One such preference, which we call *strong reciprocity* (Gintis 2000, Bowles and Gintis 2004a, Gintis, Bowles, Boyd and Fehr 2004), is a predisposition to cooperate with others, and to punish those who violate the norms of cooperation, at personal cost, even when it is implausible to expect that these costs will be repaid either by others or at a later date.

We here present empirical evidence supporting strong reciprocity as a schema for explaining important forms of political behavior. Although most of the evidence we report is based on behavioral experiments, the same behaviors are regularly observed in everyday life, for example in collective actions such as strikes and insurgencies (Petersen 2002, Goodwin, Polletta and Jasper 2001, Wood 2003), wage setting by firms (Bewley 2000), tax compliance (Andreoni, Erard and Feinstein 1998), and cooperation in the protection of local environmental public goods (Acheson 1988, Ostrom 1998, Cardenas, Stranlund and Willis 2000, Ostrom, Dietz, Dolsak, Stern, Stonich and Weber 2002).

Nothing in the material to be presented casts doubt on the rational actor framework *per se*. Our concerns address the nature and origins of preferences, not the underlying model of consequentialist choice. Decision theory shows that as long as agents have consistent and complete preferences (meaning that an agent who prefers A to B and prefers B to C also prefers A to C, and any two possible choices can be compared in terms of desirability) over a finite choice set, their actions can be modeled as if maximizing a preference function subject to constraints (Kreps 1988). Studies show that other-regarding preferences fit this framework just as well as the standard selfish preferences of traditional economic theory (Andreoni and Miller 2002). Contrary to a common usage, the fact that an action is other-regarding does not make it "irrational" or even "non-rational."

The reasons for the power of the rational actor model are clear. An agent's preferences, together with the agent's beliefs concerning the means of achieving them and the informational, material, and other constraints the agent faces have proven remarkably illuminating in accounting for individual actions. Beliefs are an indi-

vidual's conception of the relationship between an act and an outcome. Preferences are reasons for goal-oriented behavior. Preferences thus include a heterogeneous melange: tastes (food likes and dislikes, for example), habits, emotions (such as shame or anger) and other visceral reactions (such as fear), the manner in which individuals construe situations (or more narrowly, the way they frame a decision), commitments (like promises), socially enforced norms, psychological propensities (for aggression, extroversion and the like), and one's affective relationships with others. To say that a person acts on her preferences means only that knowledge of the preferences would be helpful in providing a convincing account of the actions—though not necessarily the account which would be given by the actor, for as is well known individuals are sometimes unable or unwilling to provide such an account.

We diverge from the standard preferences-beliefs-constraints model only by positing the importance of other-regarding and process-regarding behavior in accounting for human behavior in strategic interaction, and in taking the preferences accounting for this behavior as endogenous.

2 Strong Reciprocity in the Labor Market

We begin with an example of economic behavior in experimental labor markets, as it neatly illustrates the kind of motives that are present in any kind of patron client relationship or social exchange (Blau 1964). In Fehr, Gächter and Kirchsteiger (1997), the experimenters divided a group of 141 subjects (college students who had agreed to participate in order to earn money) into a set of "employers" and a larger set of "employees." The rules of the game are as follows. If an employer hires an employee who provides effort e and receives a wage w, they employer's payoff π is 100 times the effort e, minus the wage w that he must pay the employee ($\pi = 100e - w$), where the wage is between zero and 100 ($0 \leq w \leq 100$), and the effort between 0.1 and 1 ($0.1 \leq e \leq 1$). The payoff u to the employee is then the wage he receives, minus a "cost of effort," $c(e)$ ($u = w - c(e)$). The cost of effort schedule $c(e)$ is constructed by the experimenters such that supplying effort $e = 0.1$, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0 cost the employee $c(e) = 0$, 1, 2, 4, 6, 8, 10, 12, 15, and 18, respectively. All payoffs are converted into real money that the subjects are paid at the end of the experimental session.

The sequence of actions is as follows. The employer first offers a "contract" specifying a wage w and a desired amount of effort e^*. A contract is made with the first employee who agrees to these terms. An employer can make a contract (w, e^*) with at most one employee. The employee who agrees to these terms receives the wage w and supplies an effort level e, which *need not equal the contracted effort*, e^*. In effect, there is no penalty if the employee does not keep his promise, so

the employee can choose any effort level, $e \in [0.1, 1]$, with impunity. Although subjects may play this game several times with different partners, each employer-employee interaction is a one-shot (non-repeated) event. Moreover, the identity of the interacting partners is never revealed.

If employees are self-regarding, they will choose the zero-cost effort level, $e = 0.1$, no matter what wage is offered them. Knowing this, employers will never pay more than the minimum necessary to get the employee to accept a contract, which is 1 (assuming only integral wage offers are permitted). The employee will accept this offer, and will set $e = 0.1$. Since $c(0.1) = 0$, the employee's payoff is $u = 1$. The employer's payoff is $\pi = 0.1 \times 100 - 1 = 9$.

In fact, however, this self-regarding outcome rarely occurred in this experiment. The average net payoff to employees was $u = 35$, and the more generous the employer's wage offer to the employee, the higher the effort provided. In effect, employers presumed the strong reciprocity predispositions of the employees, making quite generous wage offers and receive higher effort, as a means to increase both their own and the employee's payoff, as depicted in Figure 1. Similar results have been observed in Fehr, Kirchsteiger and Riedl (1993,1998).

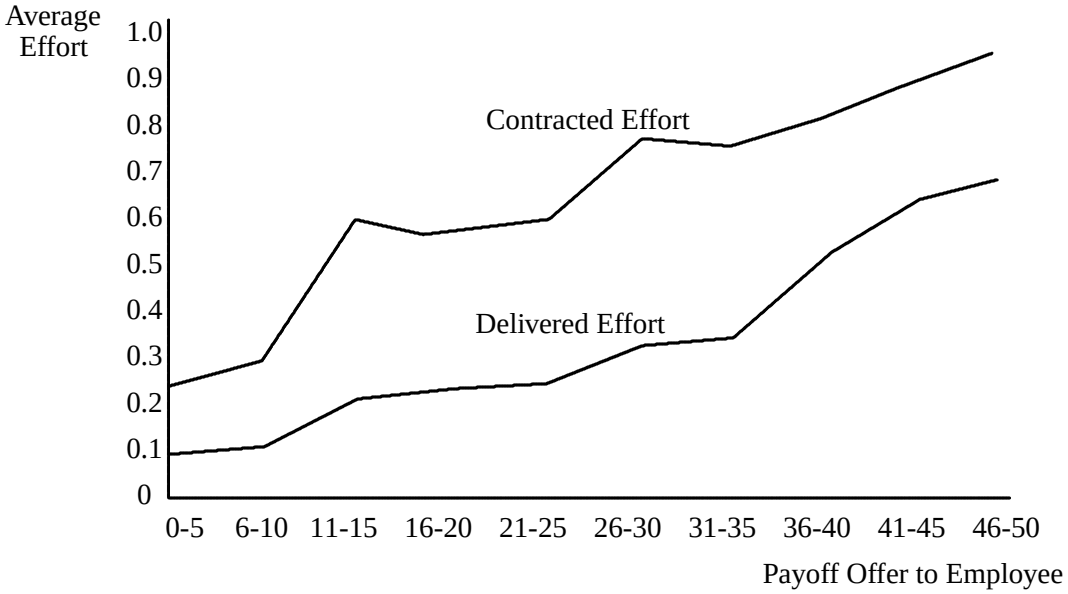

Figure 1: Relation of Contracted and Delivered Effort to Worker Payoff (141 subjects). From Fehr, Gächter, and Kirchsteiger (1997).

Figure 1 also shows that, though most employees are strong reciprocators, at any wage rate there still is a significant gap between the amount of effort agreed upon and the amount actually delivered. This is not because there are a few "bad

apples" among the set of employees, but because only 26% of employees delivered the level of effort they promised! We conclude that strong reciprocators are inclined to compromise their morality to some extent, just as we might expect from daily experience.

The above evidence is compatible with the notion that the employers are purely self-regarding, since their beneficent behavior *vis-á-vis* their employees was effective in increasing employer profits. To see if employers are also strong reciprocators, following this round of experiments, the authors extended the game by allowing the employers to respond reciprocally to the *actual effort choices* of their workers. At a cost of 1, an employer could *increase* or *decrease* his employee's payoff by 2.5. If employers were self-regarding, they would of course do neither, since they would not interact with the same worker a second time. However, 68% of the time, employers punished employees that did not fulfill their contracts, and 70% of the time, employers rewarded employees who overfulfilled their contracts. Indeed, employers rewarded 41% of employees who *exactly* fulfilled their contracts. Moreover, employees *expected* this behavior on the part of their employers, as shown by the fact that their effort levels *increased significantly* when their bosses gained the power to punish and reward them. Underfulfilling contracts dropped from 83% to 26% of the exchanges, and overfulfilled contracts rose from 3% to 38% of the total. Finally, allowing employers to reward and punish led to a 40% increase in the net payoffs to all subjects, even when the payoff reductions resulting from employer punishment of employees are taken into account. Several researchers have predicted this general behavior on the basis of general real-life social observation and field studies, including Homans (1961), Blau (1964), and Akerlof (1982). The laboratory results show that this behavior has a motivational basis in strong reciprocity and not simply long-term material self-interest.

We conclude from this study that the subjects who assume the role of "employee" conform to internalized standards of reciprocity, even when they know there are no material repercussions from behaving in a self-regarding manner. Moreover, subjects who assume the role of "employer" expect this behavior and are rewarded for acting accordingly. Finally, "employers" draw upon the internalized norm of rewarding good and punishing bad behavior when they are permitted to punish, and "employees" expect this behavior and adjust their own effort levels accordingly.

3 A Predisposition for Fairness in the Ultimatum Game

The next set of experiments evokes themes raised by Barrington Moore, Jr. (1978) in his study of obedience and revolt and James Scott (1976) in his study of rebellion in a moral economy: commitments to justice run deep, and violations of fair

treatment are likely to be harshly treated. In the ultimatum game, under conditions of anonymity, two players are shown a sum of money, say $10. One of the players, called the "proposer," is instructed to offer any number of dollars, from $1 to $10, to the second player, who is called the "responder." The proposer can make only one offer. The responder, again under conditions of anonymity, can either accept or reject this offer. If the responder accepts the offer, the money is shared accordingly. If the responder rejects the offer, both players receive nothing.

Since the game is played only once and the players do not know each other's identity, a self-regarding responder will accept any positive amount of money. Knowing this, a self-regarding proposer will offer the minimum possible amount, $1, and this will be accepted. However, when actually played, *the self-regarding outcome is never attained and never 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 generally being the modal offer), and respondents frequently reject offers below 30% (Camerer and Thaler 1995, Güth and Tietz 1990, Roth, Prasnikar, Okuno-Fujiwara and Zamir 1991).

The ultimatum game has been played around the world, but mostly with university students. We find a great deal of individual variability. For instance, in all of the above experiments a significant fraction of subjects (about a quarter, typically) behave in a self-regarding manner. But, among student subjects, average performance is strikingly uniform from country to country.

To expand the diversity of cultural and economic circumstances of experimental subjects, Henrich, Boyd, Bowles, Camerer, Fehr, Gintis and McElreath (2001) undertook a large cross-cultural study of behavior in various games including the ultimatum game. Twelve experienced field researchers, working in twelve countries on four continents, recruited subjects from fifteen small-scale societies exhibiting a wide variety of economic and cultural conditions. These societies consisted of three foraging groups (the Hadza of East Africa, the Au and Gnau of Papua New Guinea, and the Lamalera of Indonesia), six slash-and-burn horticulturists (the Aché, Machiguenga, Quichua, and Achuar of South America, and the Tsimané and Orma of East Africa), four nomadic herding groups (the Turguud, Mongols, and Kazakhs of Central Asia, and the Sangu of East Africa) and two sedentary, small-scale agricultural societies (the Mapuche of South America and Zimbabwe farmers in Africa).

We can summarize our results as follows.

The canonical model of self-regarding behavior is not supported in *any* society studied. In the ultimatum game, for example, in all societies either respondents, or proposers, or both, behaved in a reciprocal manner.

There is considerably more behavioral variability across groups than had been

found in previous cross-cultural research. While mean ultimatum game offers in experiments with student subjects are typically between 43% and 48%, the mean offers from proposers in our sample ranged from 26% to 58%. While modal ultimatum game offers are consistently 50% among university students, sample modes with these data ranged from 15% to 50%. In some groups rejections were extremely rare, even in the presence of very low offers, while in others, rejection rates were substantial, including frequent rejections of *hyper-fair* offers (i.e. offers above 50%). By contrast, the most common behavior for the Machiguenga was to offer zero. The mean offer was 22%. The Aché and Tsimané distributions resemble American distributions, but with very low rejection rates. The Orma and Huinca (non-Mapuche Chileans living among the Mapuche) have modal offers near the center of the distribution, but show secondary peaks at full cooperation.

Differences among societies in "market integration" *and* "cooperation in production" *explain a substantial portion of the behavioral variation between groups*: the higher the degree of market integration and the higher the payoffs to cooperation, the greater the level of cooperation and sharing in experimental games. The societies were rank-ordered in five categories—"market integration" (how often do people buy and sell, or work for a wage), "cooperation in production" (is production collective or individual), plus "anonymity" (how prevalent are anonymous roles and transactions), "privacy" (how easily can people keep their activities secret), and "complexity" (how much centralized decision-making occurs above the level of the household). Using statistical regression analysis, only the first two characteristics, market integration and cooperation in production, were significant, and they together accounted for 66% of the variation among societies in mean ultimatum game offers.

Individual-level economic and demographic variables did not explain behavior either within or across groups.

The nature and degree of cooperation and punishment in the experiments was generally consistent with economic patterns of everyday life in these societies.

In a number of cases the parallels between experimental game play and the structure of daily life were quite striking. Nor was this relationship lost on the subjects themselves. Here are some examples.

The Orma immediately recognized that the public goods game was similar to the *harambee*, a locally-initiated contribution that households make when a community decides to construct a road or school. They dubbed the experiment "the harambee game" and gave generously (mean 58% with 25% maximal contributors).

Among the Au and Gnau, many proposers offered more than half the pie, and many of these "hyper-fair" offers were rejected! This reflects the Melanesian culture of status-seeking through gift giving. Making a large gift is a bid for social dominance in everyday life in these societies, and rejecting the gift is a rejection of

being subordinate.

Among the whale hunting Lamalera, 63% of the proposers in the ultimatum game divided the pie equally, and most of those who did not, offered more than 50% (the mean offer was 57%). In real life, a large catch, always the product of cooperation among many individual whalers, is meticulously divided into pre-designated parts and carefully distributed among the members of the community.

Among the Aché, 79% of proposers offered either 40% or 50%, and 16% offered more than 50%, with no rejected offers. In daily life, the Aché regularly share meat, which is being distributed equally among all other households, irrespective of which hunter made the kill.

The Hadza, unlike the Aché, made low offers and had high rejection rates in the ultimatum game. This reflects the tendency of these small-scale foragers to share meat, but with a high level of conflict and frequent attempts of hunters to hide their catch from the group.

Both the Machiguenga and Tsimané made low ultimatum game offers, and there were virtually no rejections. These groups exhibit little cooperation, exchange or sharing beyond the family unit. Ethnographically, both show little fear of social sanctions and care little about "public opinion."

The Mapuche's social relations are characterized by mutual suspicion, envy, and fear of being envied. This pattern is consistent with the Mapuche's post-game interviews in the ultimatum game. Mapuche proposers rarely claimed that their offers were influenced by fairness, but rather by a fear of rejection. Even proposers who made hyper-fair offers claimed that they feared rare spiteful responders, who would be willing to reject even 50/50 offers.

4 Cooperation and Altruistic Punishment in the Public Goods Game

Our final set of experiments illuminates the tension between free riding and civic virtue central to the master works of political theory since Hume and Rousseau. The *public goods game* has been analyzed in a series of papers by the social psychologist Toshio Yamagishi (1986,1988), by the political scientist Elinor Ostrom and her coworkers (Ostrom, Walker and Gardner 1992), and by economists Ernst Fehr and his coworkers (Gächter and Fehr 1999, Fehr and Gächter 2000,2002). These researchers uniformly found that *groups exhibit a much higher rate of cooperation than can be expected assuming the standard economic model of the self-regarding actor*, and this is especially the case when subjects are given the option of incurring a cost to themselves in order to punish free riders.

A typical public goods game consists of a number of rounds, say ten. The

subjects are told the total number of rounds, as well as all other aspects of the game. The subjects are paid their winnings in real money at the end of the session. In each round, each subject is grouped with several other subjects—say 3 others—under conditions of strict anonymity. Each subject is then given a certain number of 'points,' say twenty, redeemable at the end of the experimental session for real money. Each subject then places some fraction of his points in a 'common account,' and the remainder in the subject's 'private account.' The experimenter then tells the subjects how many points were contributed to the common account, and adds to the private account of each subject some fraction, say 40%, of the total amount in the common account. So if a subject contributes his whole twenty points to the common account, each of the four group members will receive eight points at the end of the round. In effect, by putting the whole endowment into the common account, a player loses twelve points but the other three group members gain in total 24 ($= 8 \times 3$) points. The players keep whatever is in their private account at the end of the round.

A self-regarding player will contribute nothing to the common account. However, only a fraction of subjects in fact conform to the self-interest model. Subjects begin by contributing on average about half of their endowment to the public account. The level of contributions decays over the course of the ten rounds, until in the final rounds most players are behaving in a self-regarding manner (Dawes and Thaler 1988, Ledyard 1995). In a meta-study of twelve public goods experiments Fehr and Schmidt (1999) found that in the early rounds, average and median contribution levels ranged from 40% to 60% of the endowment, but in the final period 73% of all individuals ($N = 1042$) contributed nothing, and many of the remaining players contributed close to zero. These results are not compatible with the self-regarding actor model, which predicts zero contribution on all rounds, though they might be predicted by a reciprocal altruism model, since the chance to reciprocate declines as the end of the experiment approaches. However this is not in fact the explanation of moderate but deteriorating levels of cooperation in the public goods game.

The explanation of the decay of cooperation offered by subjects when debriefed after the experiment is that cooperative subjects became angry at others who contributed less than themselves, and retaliated against free-riding low contributors in the only way available to them—by lowering their own contributions (Andreoni 1995).

Experimental evidence supports this interpretation. When subjects are allowed to punish noncontributors, they do so at a cost to themselves (Orbell, Dawes, and Van de Kragt, 1986; Sato 1987; Yamagishi, 1988a,1988b,1992). For instance, in Ostrom et al. (1992) subjects interacted for twenty-five periods in a public goods 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, the only Nash equilibrium in this game that does not depend on incredible threats is for no player to pay the fee, so no player is ever punished for defecting, and all players defect by contributing nothing to the common pool. However the authors found a significant level of punishing behavior.

These studies allowed individuals to engage in strategic behavior, since costly punishment of defectors could increase cooperation in future periods, yielding a positive net return for the punisher. Fehr and Gächter (2000) set up an experimental situation in which *the possibility of strategic punishment was removed*. They used six and ten round public goods games with groups of size four, and with costly punishment allowed at the end of each round, employing three different methods of assigning members to groups. There were sufficient subjects to run between 10 and 18 groups simultaneously. Under the *Partner* treatment, the four subjects remained in the same group for all ten periods. Under the *Stranger* treatment, the subjects were randomly reassigned after each round. Finally, under the *Perfect Stranger* treatment the subjects were randomly reassigned and assured that they would never meet the same subject more than once. Subjects earned an average of about $35 for an experimental session.

Fehr and Gächter (2000) performed their experiment for ten rounds with punishment and ten rounds without (for additional experimental results and their analysis, see Bowles and Gintis (2002) and Fehr and Gächter (2002).) Their results are illustrated in Figure 2. We see that when costly punishment is permitted, cooperation does not deteriorate, and in the Partner game, despite strict anonymity, cooperation increases almost to full cooperation, even on the final round. When punishment is not permitted, however, the same subjects experience the deterioration of cooperation found in previous public goods games. The contrast in cooperation rates between the Partner and the two Stranger treatments is worth noting, because the strength of punishment is roughly the same across all treatments. This suggests that the credibility of the punishment threat is greater in the Partner treatment because in this treatment the punished subjects are certain that, once they have been punished in previous rounds, the punishing subjects are in their group. The prosociality impact of strong reciprocity on cooperation is thus more strongly manifested, the more coherent and permanent the group in question.

5 Conclusion

The evidence for other-regarding, process-regarding, and endogenous preferences is compelling. But, it raises a puzzle, one that we address in greater detail in

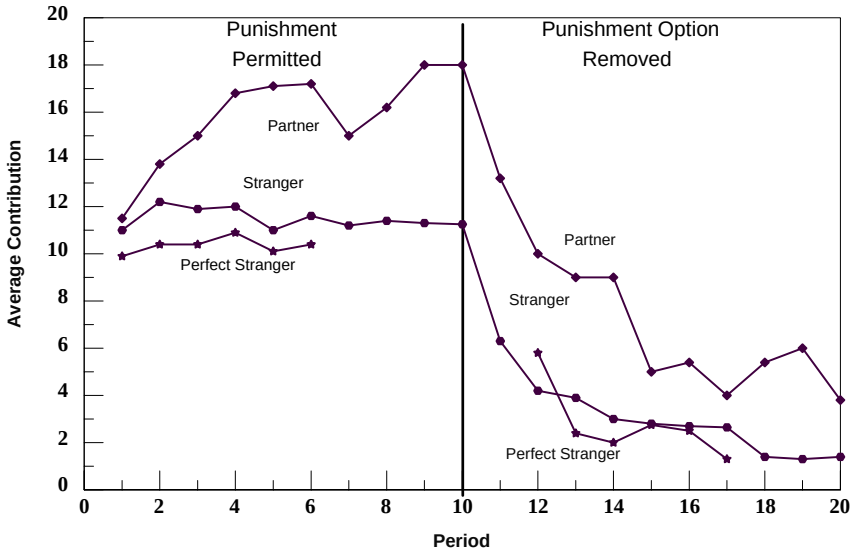

Figure 2: Average Contributions over Time in the Partner, Stranger, and Perfect Stranger Treatments when the Punishment Condition is Played First (adapted from Fehr and Gächter, 2000).

a related paper (Bowles and Gintis 2005). If many of us are fair-minded and reciprocal, then we must have acquired these preferences somehow, and it would be a good check on the plausibility of the views advanced here and the empirical evidence on which they are based to see if a reasonable account of the evolutionary success of these preferences can be provided. Generosity toward one's biological kin is readily explained (Hamilton 1964). The evolutionary puzzle concerns non-selfish behaviors towards non-kin. Among non-kin, selfish preferences would seem to be favored by any payoff-rewarding evolutionary process, whether genetic or cultural. Thus, the fair-mindedness that induces people to transfer resources to the less well-off, and the reciprocity motives that impel us to incur the costs of punishing those who violate group norms, on this account, are doomed to extinction by long term evolutionary processes. If other regarding preferences are common, this conventional evolutionary account must be incorrect.

In many cases, the evolutionary success of what appear to be unselfish traits is explained by the fact that when an accounting of long-term and indirect effects is done, the behaviors are payoff maximizing, often representing forms of mutualism. The great hunter who shares his prey may, by advertising his prowess, recruit coalition

partners and mates and deter opponents (Gintis, Smith and Bowles 2001). But, some seemingly generous behaviors are just what they seem. Indeed, the experiments we have cited were designed to study behavior in the *absence* of the indirect or long-term benefits just mentioned. The behaviors observed in these experiments, we think, have become common because they contribute to the success of groups in which the behaviors are common. People in successful groups tend to be copied, either genetically or culturally, and thus genuinely other-regarding preferences can proliferate. Recent theoretical modeling, anthropological studies, and agent-based computer simulations lend some credibility to this account (Gintis 2000, Boehm 2000, Bowles, Choi and Hopfensitz 2003, Gintis et al. 2004, Bowles and Gintis 2004b, Bowles and Gintis 2004a)

The experimental evidence as well as observation of economic and political behavior in natural settings does not lead us to reject the rational actor model, for that model, in its minimalist conception as consistency and completeness of preferences, is perfectly compatible with altruistic, spiteful, or reciprocal motives. Indeed, this versatility is among its merits.

However, an adequate reformulation of the psychological foundations of the behavioral sciences cannot be accomplished by inventing some new *Homo sociologicus* or *zoon politikon* to replace *Homo economicus* as the epitome of intentional behavior. Behavioral experiments and everyday observation make it clear that populations are heterogeneous. Heterogeneity makes a difference in outcomes. But, as the public goods experiments showed, its effects are not adequately captured by a process of simple averaging. The outcome of interaction among a population that is composed of equal numbers of saints and sinners will not generally be the average of the outcomes of two populations with just one type. The reason is that in many settings, the norm-upholding activities of a few saints may induce even the sinners to act civic-mindedly, while in other institutional settings, a few sinners can induce all players to act like *Homo economicus*. Recall, as another example, that in the public goods-with-punishment game, those with reciprocal preferences not only acted generously themselves, but they apparently also induced the selfish types to act as if they were generous.

Indeed, seemingly small differences in institutions can make large differences in outcomes, as illustrated by the following example. Imagine a one-shot Prisoners' Dilemma game played between a self-regarding player, for whom defect is the dominant strategy in the simultaneous moves game, and a strong reciprocator, who prefers to cooperate if the other cooperates and to defect otherwise (Kiyonari, Tanida and Yamagishi 2000, Fehr and Fischbacher 2001). Suppose the players' types are known to each. If the game is played simultaneously, the reciprocator, knowing that the other will defect, will do the same. The outcome will be mutual defection. If the self-regarding player moves first, however, he will know that the reciprocator will

match whatever action he takes, narrowing the possible outcomes to {cooperate, cooperate} or {defect, defect}, the former yielding both players a higher payoff. The self-regarding first mover will therefore cooperate and mutual cooperation will be sustained as the outcome.

In addition to heterogeneity across individuals, versatility of individuals must also be accounted for. In the ultimatum game, many proposers often offer amounts that maximize their expected payoffs, given the observed relationship between offers and rejections: they behave selfishly but expect responders not to. And they are correct in this belief! The *same individuals*, when in the role of responder, typically reject substantial offers if they appear to be unfair, thus confirming the expectations of the proposer and violating the self-interest axiom.

Finally, as our cross-cultural experiments suggest, culture matters: differences in an individual's preferences often correspond to differences in the way people interact socially in making their living and in other aspects of daily life. This means that populations that experience different structures of social interaction over prolonged periods are likely to exhibit differing behaviors, not simply because the constraints entailed by these institutions are different but also because the structure of social interaction affects the evolution of preferences.

Progress in the direction of a more adequate behavioral foundation for political behavior must take account of these three aspects of people: namely their *heterogeneity*, their *versatility*, and their *plasticity*.

REFERENCES

Acheson, J., *The Lobster Gangs of Maine* (Hanover, NH: New England Universities Press, 1988).

Akerlof, George A., "Labor Contracts as Partial Gift Exchange," *Quarterly Journal of Economics* 97,4 (November 1982):543–569.

Andreoni, James, "Cooperation in Public Goods Experiments: Kindness or Confusion," *American Economic Review* 85,4 (1995):891–904.

— and John H. Miller, "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism," *Econometrica* 70,2 (2002):737–753.

— , Brian Erard, and Jonathan Feinstein, "Tax Compliance," *Journal of Economic Literature* 36,2 (June 1998):818–860.

Bewley, Truman F., *Why Wages Don't Fall During a Recession* (Cambridge: Cambridge University Press, 2000).

Blau, Peter, *Exchange and Power in Social Life* (New York: John Wiley, 1964).

Boehm, Christopher, *Hierarchy in the Forest: The Evolution of Egalitarian Behavior* (Cambridge, MA: Harvard University Press, 2000).

Bowles, Samuel and Herbert Gintis, "Homo Reciprocans," *Nature* 415 (10 January 2002):125–128.

— and — , "The Evolution of Strong Reciprocity: Cooperation in Heterogeneous Populations," *Theoretical Population Biology* 65 (2004):17–28.

— and — , "The Origins of Human Cooperation," in Peter Hammerstein (ed.) *Genetic and Cultural Origins of Cooperation* (Cambridge, MA: MIT Press, 2004).

— and — , "Evolutionary Origins of Collective Action," in Donald Wittman and Barry Weingast (eds.) *Oxford Handbook of Political Economy* (Oxford: Oxford University Press, 2005).

— , Jung-kyoo Choi, and Astrid Hopfensitz, "The Co-evolution of Individual Behaviors and Social Institutions," *Journal of Theoretical Biology* 223 (2003):135–147.

Camerer, Colin and Richard Thaler, "Ultimatums, Dictators, and Manners," *Journal of Economic Perspectives* 9,2 (1995):209–219.

Cardenas, Juan Camilo, John K. Stranlund, and Cleve E. Willis, "Local Environmental Control and Institutional Crowding-out," *World Development* 28,10 (July 2000):1719–1733.

Dawes, Robyn M. and Richard Thaler, "Cooperation," *Journal of Economic Perspectives* 2 (1988):187–197.

Fehr, Ernst and Klaus M. Schmidt, "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics* 114 (August 1999):817–868.

— and Simon Gächter, "Cooperation and Punishment," *American Economic Review* 90,4 (September 2000):980–994.

— and — , "Altruistic Punishment in Humans," *Nature* 415 (10 January 2002):137–140.

— and Urs Fischbacher, "Why Social Preferences Matter," 2001. Nobel Symposium on Behavioral and Experimental Economics.

— , Georg Kirchsteiger, and Arno Riedl, "Does Fairness Prevent Market Clearing?," *Quarterly Journal of Economics* 108,2 (1993):437–459.

— , — , and — , "Gift Exchange and Reciprocity in Competitive Experimental Markets," *European Economic Review* 42,1 (1998):1–34.

— , Simon Gächter, and Georg Kirchsteiger, "Reciprocity as a Contract Enforcement Device: Experimental Evidence," *Econometrica* 65,4 (July 1997):833–860.

Gächter, Simon and Ernst Fehr, "Collective Action as a Social Exchange," *Journal of Economic Behavior and Organization* 39,4 (July 1999):341–369.

Gintis, Herbert, "Strong Reciprocity and Human Sociality," *Journal of Theoretical Biology* 206 (2000):169–179.

— , Eric Alden Smith, and Samuel Bowles, "Costly Signaling and Cooperation," *Journal of Theoretical Biology* 213 (2001):103–119.

— , Samuel Bowles, Robert Boyd, and Ernst Fehr, *Moral Sentiments and Material Interests: On the Foundations of Cooperation in Economic Life* (MIT Press, 2004).

Goodwin, Jeff, Francesca Polletta, and James. M. Jasper, *Passionate Politics: Emotions and Social Movements* (Chicago: University of Chicago Press, 2001).

Güth, Werner and Reinhard Tietz, "Ultimatum Bargaining Behavior: A Survey and Comparison of Experimental Results," *Journal of Economic Psychology* 11 (1990):417–449.

Hamilton, William D., "The Genetical Evolution of Social Behavior I & II," *Journal of Theoretical Biology* 7 (1964):1–16,17–52.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath, "Cooperation, Reciprocity and Punishment in Fifteen Small-scale Societies," *American Economic Review* 91 (May 2001):73–78.

Homans, George, *Social Behavior: Its Elementary Forms* (New York: Harcourt Brace, 1961).

Kiyonari, Toko, Shigehito Tanida, and Toshio Yamagishi, "Social Exchange and Reciprocity: Confusion or a Heuristic?," *Evolution and Human Behavior* 21 (2000):411–427.

Kreps, David M., *Notes on the Theory of Choice* (London: Westview, 1988).

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

Moore, Jr., Barrington, *Injustice: The Social Bases of Obedience and Revolt* (White Plains: M. E. Sharpe, 1978).

Orbell, John M., Robyn M. Dawes, and J. C. Van de Kragt, "Organizing Groups for Collective Action," *American Political Science Review* 80 (December 1986):1171–1185.

Ostrom, Elinor, "A Behavioral Approach to the Rational Choice Theory of Collective Action," *American Political Science Review* 92,1 (March 1998):1–21.

— , James Walker, and Roy Gardner, "Covenants with and without a Sword: Self-Governance Is Possible," *American Political Science Review* 86,2 (June 1992):404–417.

— , Thomas Dietz, Nives Dolsak, Paul C. Stern, Susan Stonich, and Elke Weber, *The Drama of the Commons* (Washington: National Academy Press, 2002).

Petersen, Roger D., *Understaneing Ethnic Violence, Fear, Hatred, and Resentment in Twentieth-century Eastern Europe* (Cambridge: Cambridge University Press, 2002).

Roth, Alvin E., Vesna Prasnikar, Masahiro Okuno-Fujiwara, and Shmuel Zamir, "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study," *American Economic Review* 81,5 (December 1991):1068–1095.

Sato, Kaori, "Distribution and the Cost of Maintaining Common Property Resources," *Journal of Experimental Social Psychology* 23 (January 1987):19–31.

Scott, James C., *The Moral Economy of the Peasant: Rebellion and Subsistence in Southeast Asia* (New Haven, CT: Yale University Press, 1976).

Wood, Elisabeth Jean, *Insurgent Collective Action and Civil War in El Salvador* (Cambridge,: Cambridge University Press, 2003).

Yamagishi, Toshio, "The Provision of a Sanctioning System as a Public Good," *Journal of Personality and Social Psychology* 51 (1986):110–116.

— , "The Provision of a Sanctioning System in the United States and Japan," *Social Psychology Quarterly* 51,3 (1988):265–271.

— , "Seriousness of Social Dilemmas and the Provision of a Sanctioning System," *Social Psychology Quarterly* 51,1 (1988):32–42.

— , "Group Size and the Provision of a Sanctioning System in a Social Dilemma," in W.B.G. Liebrand, David M. Messick, and H.A.M. Wilke (eds.) *Social Dilemmas: Theoretical Issues and Research Findings* (Oxford: Pergamon Press, 1992) pp. 267–287.