The Evolutionary Basis of Collective Action

Samuel Bowles
Herbert Gintis*

April 20, 2006

1 Introduction

Many aspects of political behavior have been illuminated by standard models in which political actors maximize self-interested preferences. The works of Downs (1957), Buchanan and Tullock (1962),Buchanan, Tollison and Tullock (1980) and Becker (1983), as well as those inspired by these seminal contributions, have contributed to our understanding of voter, party and policy preferences, interest group politics, rent-seeking, coalition formation, bargaining and other aspects of political behavior. Using this framework, works on electoral support for the welfare state (Benabou and Ok 2001, Moene and Wallerstein 2002), informal enforcement of contracts (Greif 1994, Greif, Milgrom and Weingast 1994), the efficiency of democratic governance (Wittman 1989), nationalism (Breton, Galeotti, Salmon and Wintrobe 1995) and ethnic conflict (Varshney 2003) have produced important and sometimes surprising insights.

Yet as Ostrom (1998) and others have pointed out, a number of critical aspects of political behavior remain difficult to explain within this framework. These include the fact that people bother to vote at all, and electoral support for costly redistributive programs from which the voter concerned is unlikely to benefit and for which he will certainly pay additional taxes (Luttmer 2001, Fong 2001, Fong, Bowles and Gintis 2005), and many forms of political violence (Stern 2003). Among the more striking examples of the shortcomings of the standard model is the large class of political behavior that takes the form of voluntary contribution to public goods. Included is participation in joint political activities and other forms of collective

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

action (Moore, Jr. 1978, Wood 2003, Scott 1976), the adherence to social norms (Young and Burke 2001, Andreoni, Erard and Feinstein 1998), and the punishment of those violating social norms (Mahdi 1986, Harding 1978, Boehm 1993, Wiessner 2005).

When one is motivated to bear personal costs to help or to hurt others we say that one has *other-regarding* preferences, meaning that affecting the states experienced by someone other than oneself is part of one's motivations. Unlike the conventional self-regarding preferences of *Homo economicus*, social preferences are other regarding. Generosity towards others, and punishing those who violate norms are commonly motivated by other-regarding preferences.

We use the term self-regarding rather than "selfish" to describe the standard assumptions about preferences to avoid the circularity arising from the fact that all uncoerced actions are motivated by preferences and hence might confusingly be termed selfish, leaving only those actions that violate one's preference ordering to be called unselfish (but would better be called non-rational). To explain behavior, both other-regarding and self regarding preferences must be transitive, and when they are (as we assume) the actions they motivate are rational in the strict sense typically adopted in economics and decision theory. The common designation of generous behavior as "irrational" is based on a gratuitous conflation of rationality and self-regarding preferences.

We explore two problems in the study of the political behaviors supporting collective action. The first concerns the view frequently advanced by economists and biologists that cooperative behaviors can be fully explained on the basis of self-interested motivations, once one takes account of the repeated nature of interactions and the degree of genetic relatedness among members of a cooperating group. We show that repeated interactions and kin-based altruism, while strong influences on behavior in many settings, do not provide an adequate account of the forms of cooperation observed in natural and experimental settings.

These and other types of political behavior are based on preferences that include a concern for the well being of others and a taste not only for fairness but also for retribution. We review recent behavioral experiments documenting the variety and extent of these so called social preferences and the manner in which the existence of even a minority of individuals with social preferences can dramatically affect group behavior (see Bowles and Gintis (2007b), Gintis, Bowles, Boyd and Fehr (2005), and Henrich, Boyd, Bowles, Camerer, Fehr and Gintis (2004) for a more extensive review of this evidence).

The second is the puzzle of how these social preferences could have evolved by means of genetic transmission and natural selection, cultural learning and socialization, or both. The puzzle arises because the political behaviors motivated by social preferences are often altruistic in the biological sense—of conferring gains

on others in one's group while entailing costs—and altruistic behaviors will be disadvantaged in most evolutionary processes that favor higher payoff types. Our treatment of these topics is necessarily cursory, drawing extensively on work presented more fully in Bowles and Gintis (2007a), Gintis et al. (2005), and Henrich et al. (2004).

2 The Cooperative Species

Cooperation among humans is unique in nature, extending to a large number of unrelated individuals and taking a vast array of forms. By cooperation we mean engaging with others in a mutually beneficial activity. Cooperative behavior may confer benefits net of costs on the individual cooperator, and thus may be motivated by entirely self-regarding preferences. In this case, cooperation is a form of what biologists call *mutualism*, namely an activity that confers net benefits both on the actor and on others.

But, cooperation may also incur net costs to the individual. In this case cooperative behavior constitutes a form of *altruism*. In contrast to mutualistic cooperation, altruistic cooperation would not be undertaken by an individual whose motives were entirely self-regarding and thus did not take account of the effects of one's actions on others.

While the high frequency of altruistic cooperation in humans relative to other species could be an evolutionary accident, a more plausible explanation is that altruistic cooperation among humans is the result of capacities that are unique to our species and that strongly promote our relative reproductive fitness. Thus, we seek an explanation of cooperation that works for humans, but which, because it involves capacities that are unique to humans, does not work for other species, or works substantially less well.

Central to our explanation will be human cognitive, linguistic and physical capacities that allow the formulation of general norms of social conduct, the emergence of social institutions regulating this conduct, the psychological capacity to internalize norms, and the capacity to base group membership on such non-kin characteristics as ethnicity and linguistic differences, which in turn facilitates costly conflicts among groups. Also important is the unique human capacity to use projectile weapons, a consequence of which is to lower the cost of punishing norm violators within a group, and to render intergroup conflicts more lethal.

Thus, our account of human sociality and its evolution hinges critically on a reconsideration of the canonical economic model of self-interested behavior. But more than individual motivation is involved. The extraordinary levels of cooperation observed in human society cannot be attributed simply to our generosity towards

those with whom we interact or our capacity to favor the advancement of our nation or ethnic group over our individual well-being. The regulation of social interactions by group-level norms and institutions plays no less a role than altruistic individual motives in understanding how the cooperative species came to be. The institutions that regulate behaviors among non-kin affect the rewards and penalties associated with particular behaviors, often favoring the adoption of cooperative actions over others. In the social environments common to human interactions, the self-regarding are often induced to act in the interest of the group. Of course it will not do to posit these rules and institutions *a priori*. Rather, we show that these could have co-evolved with other human traits in a plausible representation of the relevant ecologies and social environments.

Cooperation is not an end to be valued in its own right, but rather is a means that under some conditions may contribute to human well being. In other settings, competition plays no less essential a role. Similarly, the individual motives and group-level institutions that account for cooperation among humans include not only the most elevated—a concern for others, fair-mindedness, and democratic accountability of leaders, for example—but also the most venal: vengeance, exclusion of "outsiders," and frequent warfare among groups, for example.

Our reasoning is disciplined in three ways. First, the forms of cooperation we seek to explain are confirmed by natural observation, historical accounts, and behavioral experiments. Second, our account is based on a plausible evolutionary dynamic involving some combination of genetic and cultural transmission, the consistency of which can be demonstrated through formal modeling. Third, agent-based simulations show that our models can account for human cooperation under parameter values consistent with what can be reasonably inferred about the environments in which humans evolved.

3 Mutualistic Cooperation

Because mutualistic cooperation will be sustained by individuals with entirely self-regarding preferences, it is treated in standard biological and economic models as an expression of self-interest. "Natural selection favors these…behaviors," wrote Robert Trivers in his "The Evolution of Reciprocal Altruism" (1971), "because in the long run they benefit the organism performing them.…two individuals who risk their lives to save each other will be selected over those who face drowning on their own." (pp. 34–35) Cooperation, in Trivers' interpretation is simply symbiosis with a time lag. Trivers' explanation initially found favor among biologists and economists because it is consistent with both the common biological reasoning that natural selection will not favor altruistic behaviors and with the canonical economic

assumption of self interest.

Trivers identified the conditions under which assisting another would be reciprocated in the future with a likelihood sufficient to make mutual assistance a form of mutualism. These conditions favoring reciprocal altruism included an extended lifetime, mutual dependence and other reasons for limited dispersal so that groups remain together, extended periods of parental care, attenuated dominance hierarchies, and frequent combat with conspecifics and predators. Foraging bands of humans, he pointed out, exhibit all of these conditions. Michael Taylor (1976) and Robert Axelrod and William Hamilton (1981) subsequently formalized Trivers' argument using the theory of repeated games. In economics, analogous reasoning is summarized in the Folk Theorem, which shows that cooperation among self-regarding individuals can be sustained as long as interactions are expected to be repeated with sufficient frequency and individuals are not too impatient (Fudenberg and Maskin 1986, Fudenberg, Levine and Maskin 1994).

But, in many important human social environments, Trivers' conditions favoring reciprocal altruism do not hold, yet cooperation among non-kin is commonly observed. These include contributing to common projects when community survival is threatened, and cooperation among very large numbers of people among who do not share common knowledge of one another's actions. In fact, the scope of application of the Folk Theorem is quite restricted, especially in groups of any significant size, once the problem of cooperation is posed in an evolutionary setting and account is taken of "noise" arising from mistaken behaviors and misinformation about the behaviors of others.

A plausible model of cooperation must satisfy the following five conditions. First, it must be *incentive compatible*. In particular, those who provide the rewards and inflict punishments dictated by the rules for cooperation must have the motivation to do so. Second, a model must be *dynamically stable*, in the sense that random fluctuations, errors, and mutations (the emergence of novel strategies) do not disrupt cooperation or entail excessive efficiency losses. Third, the organizational forms and incentive mechanisms deployed in the model must reflect the types of strategic interaction and incentives widely observed in human groups. In particular, the model should work well with group sizes on the order of ten to twenty, and the incentive to punish defectors should reflect those deployed in real-world public goods game settings. Fourth, the model should not require extraordinary *informational requirements*. Finally the model should work with *plausible discount factors*. It is reasonable to suppose that within a group faced by a public goods game, there will be a distribution of discount factors among members, and average discount factors can be high in some periods and low in others, as the probability of group dissolution rises and falls.

A careful analysis shows that all models of cooperation based on tit-for-tat

and related repeated game strategies when played among self-interested individuals violate these conditions, and hence fail to solve the problem of cooperation among unrelated agents.[1] First, reciprocal altruism fails when a social group is threatened with dissolution, since members who sacrifice now on behalf of group members do not have a high probability of being repaid in the (highly uncertain) future.

Second, many human interactions in the relevant evolutionary context took the form of n-person public goods games—food sharing and other co-insurance, upholding social norms among group members, information sharing, and common defense—rather than dyadic interactions. The difficulty in sustaining cooperation in public goods games by means of the standard tit-for-tat and related repeated game strategies increases exponentially with group size (Boyd and Richerson 1988, Bowles and Gintis 2007a), even if interactions are repeated with high probability. The reason is that in groups larger than two, withdrawing cooperation in response to a single defection imposes a blanket punishment on all, defectors and cooperators alike. But, targeting punishment on defectors alone does not work in large groups unless members have unrealistically accurate information about the actions taken by others.

Third, the contemporary study of human behavior has documented a large class of social behaviors inexplicable in terms of reciprocal altruism. For instance, there is extensive support for income redistribution in advanced industrial economies, even among those who cannot expect to be net beneficiaries (Fong, Bowles and Gintis, 2005). Under some circumstances group incentives for large work teams are effective motivators even when the opportunity for reciprocation is absent and the benefits of cooperation are so widely shared that a self-interested group member would gain from free-riding on the effort of others (Ghemawat 1995, Hansen 1997, Knez and Simester 2001). Finally, laboratory and field experiments show that other-regarding motives are frequently robust causes of cooperative behavior, even in one-shot, anonymous settings.

4 Strong Reciprocity: Evidence from Behavioral Experiments

A more direct reason for doubting the interpretation that most cooperation is mutualistic is given by the compelling evidence that many (perhaps most) people behave in ways inconsistent with the assumption that they are motivated by self-regarding

[1] This analysis is presented in full in Gintis (2004) and Bowles and Gintis (2007a), which also shows that recent game-theoretic extensions of these models using repeated game theory (Fudenberg and Maskin 1986, Fudenberg et al. 1994, Sekiguchi 1997, Piccione 2002, Ely and Välimäki 2002, Bhaskar and Obara 2002, Matsushima 2000, Kandori 2002) do not alter this conclusion. These contributions, while important in their own right, either suffer the same problems discussed in the text, or they are not stable in a dynamic setting.

preferences. A suggestive body of evidence points to the importance of a suite of behaviors that we call *strong reciprocity*. A strong reciprocator comes to a new social situation with a predisposition to cooperate, is predisposed to respond to cooperative behavior on the part of others by maintaining or increasing his level of cooperation, and responds to anti-social behavior on the part of others by retaliating against the offenders, even at a cost to himself, and even when he cannot not reasonably expect future personal gains from such retaliation. The strong reciprocator is thus both a *conditionally altruistic cooperator* and a *conditionally altruistic punisher* whose actions benefit other group members at a personal cost. We call this "strong reciprocity" to distinguish it from "weak" (i.e., self-regarding) forms of reciprocity, such as Trivers' reciprocal altruism.

Strong reciprocity is an example of a larger class of so-called *social preferences* which describe the motivations of people who care (one way or the other) about the well being of others, and have preferences not only over the states they and others experience but also care about how the states came about.

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-interested responder will accept any positive amount of money. Knowing this, a self-interested proposer will offer the minimum possible amount, $1, and this will be accepted. However, when actually played, *the self-interested 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).

Strong reciprocity emerges in many other experimental games, some of which are described in Table 1 (from Camerer and Fehr, 2003). In all cases, given the one-shot, anonymous nature of the game, self-regarding agents would neither contribute to the common good, or reward others for so contributing. Nor would they punish others for failing to contribute. Yet, in each game, under many different conditions and in different cultures, a considerable fraction of agents contributes, and enough agents punish free-riding that even the self-regarding agent often contribute simply to avoid punishment.

5 The Evolution of Strong Reciprocity

If preferences were entirely self-regarding,the extent of human cooperation would indeed be puzzling. But if social preferences are common, the puzzle takes a somewhat different form: how might strong reciprocity and other altruistic preferences that support cooperation have evolved over the course of human history? The puzzle is posed especially clearly if the processes of cultural and genetic evolution favor behavioral traits that on the average are associated with higher levels of material success. We think that this assumption of what is called a *payoff monotonic dynamic* is not entirely adequate. But Gintis (2000) and Bowles and Gintis (2004) adopt just such an evolutionary model to show that individuals behaving as strong reciprocators can proliferate in a population in which they were initially rare, and that their presence in a population could sustain high levels of cooperation among group members.

One intuition behind these models is that in groups with strong reciprocators present, group members whose self-regarding preferences lead them to shirk on contributing to common projects will be punished by being ostracized from the group. Strong reciprocators bear the cost not only of contributing to common projects, but also of punishing the shirking of the self-interested members. If reciprocators are common enough, however, the self-interested members will conform to cooperative norms in order to escape punishment, thereby reducing or eliminating the fitness differences between the reciprocators and the self-interested members. A second argument supporting strong reciprocity is that groups with a sufficient fraction of strong reciprocators will be better capable to survive such group crises as war, pestilence, and adverse climatic conditions. In such situations, a group of self-regarding agents would simply disband, since each member will do better to bear the personal costs of abandoning the group rather than bearing the even heavier costs of attempting to preserve the group, most of the gains of which would accrue to other group members. Since strong reciprocators enforce cooperation without regard for the possibility of extinction, a sufficient fraction of strong reciprocators can enhance the possibility of group survival.

Group level-characteristics—such as relatively small group size, limited migration, or frequent inter-group conflicts—have co-evolved with cooperative behaviors. Cooperation is thus based in part on the distinctive capacities of humans to construct institutional environments that limit within-group competition and reduce phenotypic variation within groups, thus heightening the relative importance of between-group competition, and hence allowing individually-costly but ingroup-beneficial behaviors to coevolve with these supporting environments through a process of inter-demic selection.

The idea that the suppression of within-group competition may be a strong in-

fluence on evolutionary dynamics has been widely recognized in eusocial insects and other species. Alexander (1979), Boehm (1982) and Eibl-Eibesfeldt (1982) first applied this reasoning to human evolution, exploring the role of culturally transmitted practices that reduce phenotypic variation within groups. Group-level institutions thus are constructed environments capable of imparting distinctive direction and pace to the process of biological evolution and cultural change (Friedman and Singh 2001).

Bowles, Choi and Hopfensitz (2003) models an evolutionary dynamic along these lines. They show that intergroup conflicts may explain the evolutionary success of both altruistic forms of human sociality towards non-kin, and group-level institutional structures such as resource sharing that have emerged and diffused repeatedly in a wide variety of ecologies during the course of human history.

6 Proximate Motives: Internalized Norms and Social Emotions

An *internal norm* is a pattern of behavior enforced in part by internal sanctions, including shame and guilt. Individuals follow internal norms when they value certain behaviors for their own sake, in addition to, or despite, the effects these behaviors have on personal fitness and/or perceived well-being. The ability to internalize norms is nearly universal among humans. All successful cultures foster internal norms that enhance personal fitness, such as future-orientation, good personal hygiene, positive work habits, and control of emotions. Cultures also widely promote altruistic norms that subordinate the individual to group welfare, fostering such behaviors as bravery, honesty, fairness, willingness to cooperate, and empathy with the distress of others (Brown 1991).

If even a fraction of society internalize the norms of cooperation and punish free riders and other norm violators, a high degree of cooperation can be maintained in the long run. The puzzles are two: why do we internalize norms, and why do cultures promote cooperative behaviors? Gintis (2003) provides an evolutionary model in which the capacity to internalize norms develops because this capacity enhances individual fitness in a world in which social behavior has become too complex to be learned through personal experience alone. It is not difficult to show that if an internal norm is fitness enhancing, then for plausible patterns of socialization, the allele for internalization of norms is evolutionarily stable. This framework implements the suggestion in Simon (1990) that altruistic norms can 'hitchhike' on the general tendency of internal norms to be fitness-enhancing.

Prosocial emotions are physiological and psychological reactions that induce agents to engage in cooperative behaviors as we have defined them above. The prosocial emotions include some, such as shame, guilt, empathy, and sensitivity to

social sanction, that induce agents to undertake constructive social interactions, and others, such as the desire to punish norm violators, that reduce free riding when the prosocial emotions fail to induce sufficiently cooperative behavior in some fraction of members of the social group (Frank 1987, Hirshleifer 1987). Without the prosocial emotions we would all be sociopaths, and human society would not exist, however strong the institutions of contract, governmental law enforcement, and reputation. Sociopaths have no mental deficit except that their capacity to experience shame, guilt, empathy, and remorse is severely attenuated or absent.

Prosocial emotions function like the basic emotion, "pain," in providing guides for action that bypass the explicit cognitive optimizing process that lies at the core of the standard behavioral model in economics. Antonio Damasio (1994):173 calls these "somatic markers," that is, a bodily response that "forces attention on the negative outcome to which a given action may lead and functions as an automated alarm signal which says: Beware of danger ahead if you choose the option that leads to this outcome.…the automated signal protects you against future losses." Emotions thus contribute to the decision-making process, not simply by clouding reason, but in beneficial ways as well. Damasio continues: "suffering puts us on notice.…it increases the probability that individuals will heed pain signals and act to avert their source or correct their consequences." (p. 264)

Does shame serve a purpose similar to that of pain? If being socially devalued has fitness costs, and if the amount of shame is closely correlated with the level of these fitness costs, then the answer is affirmative. Shame, like pain, is an aversive stimulus that leads the agent experiencing it to repair the situation that led to the stimulus, and to avoid such situations in the future. Shame, like pain, replaces an involved optimization process with a simple message: whatever you did, undo it if possible, and do not do it again.

Since shame is evolutionarily selected and is costly to use, it very likely confers a selective advantage on those who experience it. Two types of selective advantage are at work here. First, shame may raise the fitness of an agent who has incomplete information (e.g., as to how fitness-reducing a particular anti-social action is), limited or imperfect information-processing capacity, and/or a tendency to undervalue costs and benefit that accrue in the future. Probably all three conditions conspire to react suboptimally to social disapprobation in the absence of shame, and shame brings us closer to the optimum. Of course the role of shame in alerting us to negative consequences in the future presupposes that society is organized to impose those costs on rule violators. The emotion of shame may have coevolved with the emotions motivating punishment of antisocial actions (the reciprocity motive in our model).

The second selective advantage to those experiencing shame arises through the effects of group competition. Where the emotion of shame is common, punishment

of antisocial actions will be particularly effective and as a result seldom used .Thus groups in which shame is common can sustain high levels of group cooperation at limited cost and will be more likely to spread through interdemic group selection (Bowles and Gintis 2004, Boyd, Gintis, Bowles and Richerson 2003). Shame thus serves as a means of economizing on costly within-group punishment.

While we think the evidence is strong that prosocial emotions account for important forms of human cooperation, there is no universally accepted model of how emotions combine with more cognitive processes to affect behaviors. Nor is there much agreement on how best to represent the prosocial emotions that support cooperative behaviors.

Bowles and Gintis (2005) considers a public goods game where subjects maximize a utility function that captures five distinct motives: personal material payoffs, one's valuation of the payoffs to others, which depend both on ones' altruism and one's degree of reciprocity, and one's sense of guilt or shame when failing to contribute one's fair share to the collective effort of the group. We have evidence of shame if players who are punished by others respond by behaving more cooperatively than is optimal for a material payoff-maximizing agent. We present indirect empirical evidence suggesting that such emotions play a role in the public goods game.

Direct evidence on the role of emotions in experimental games remains scanty. The forms of arousal associated with emotions are readily measured, but they do not readily allow us to distinguish between, say, fear and anger. Self reports of emotional states are informative but noisy. Recent advances in brain imaging, however, can identify the areas of the brain that are activated when an experimental subject is confronted with a moral dilemma or unfair treatment by another experimental subject. This use of fMRI and related technology may eventually allow us to distinguish among the emotional responses of subjects in experimental situations.

7 Conclusion

The study of collective action and other forms of cooperative behaviors exhibits a curious disparity among social scientists. In the Marxian tradition, and among many historians, sociologists, anthropologists, and political scientists, the fact that people often behave prosocially in the pursuit of common objectives, even when this involves cooperating in an n-person Prisoners Dilemma game, is frequently invoked to explain social structures and their dynamics. Among economists, biologists, and others influenced by their models, by contrast, self-regarding actors will rarely, if ever, cooperate in such a setting.

It may be thought that the key difference accounting for this divergence is the

methodological individualism adopted by economists and biologists, in contrast to the more holist or structural approaches adopted by historians and many social scientists outside of economics. According to this view, if anthropologists, sociologists, Marxists, and others were only to ask the obvious question—why would an individual engage in a costly activity to benefit others?—they would agree with the economists. But this is not the case.

The question needs an answer, but in light of what we now know about the nature of social preferences, it is not that altruistic forms of collective action are likely to be an ephemeral and unimportant aspect of political life and that most forms of seemingly altruistic cooperation are just self-interest in disguise. Like adherence to social norms and punishment of those who violate them, collective action is an essential aspect of political behavior and one which is readily explained by the fact that strong reciprocity and other social preferences are sufficiently common in most human populations to support high levels of cooperation in many social settings.

References

Alexander, R. D., *Biology and Human Affairs* (Seattle: University of Washington Press, 1979).

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

Axelrod, Robert and William D. Hamilton, "The Evolution of Cooperation," *Science* 211 (1981):1390–1396.

Becker, Gary, "A Theory of Competition among Pressure Groups for Political Influence," *Quarterly Journal of Economics* 98,3 (1983):371–400.

Benabou, Roland and Efe A. Ok, "Social Mobility and the Demand for Redistribution: The Poum Hypothesis," *Quarterly Journal of Economics* 116,2 (May 2001):447–487.

Berg, Joyce, John Dickhaut, and Kevin McCabe, "Trust, Reciprocity, and Social History," *Games and Economic Behavior* 10 (1995):122–142.

Bhaskar, V. and Ichiro Obara, "Belief-Based Equilibria the Repeated Prisoner's Dilemma with Private Monitoring," *Journal of Economic Theory* 102 (2002):40–69.

Boehm, Christopher, "The Evolutionary Development of Morality as an Effect of Dominance Behavior and Conflict Interference," *Journal of Social and Biological Structures* 5 (1982):413–421.

— , "Egalitarian Behavior and Reverse Dominance Hierarchy," *Current Anthropology* 34,3 (June 1993):227–254.

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

— and — , "Prosocial Emotions," in Lawrence E. Blume and Steven N. Durlauf (eds.) *The Economy As an Evolving Complex System III* (Santa Fe, NM: Santa Fe Institute, 2005).

— and — , "A Cooperative Species: Human Reciprocity and its Evolution," 2007. in preparation.

— and — , "Social Preferences, Homo Economicus and Zoon Politikon," in Robert Goodin and Charles Tilly (eds.) *Oxford Handbook of Contextual Political Studies* (Oxford: Oxford University Press, 2007).

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

Boyd, Robert and Peter J. Richerson, "The Evolution of Reciprocity in Sizable Groups," *Journal of Theoretical Biology* 132 (1988):337–356.

— , Herbert Gintis, Samuel Bowles, and Peter J. Richerson, "Evolution of Altruistic Punishment," *Proceedings of the National Academy of Sciences* 100,6 (March 2003):3531–3535.

Breton, A., G. Galeotti, P. Salmon, and R. Wintrobe, *Nationalism and Rationality* (Cambridge: Cambridge University Press, 1995).

Brown, Donald E., *Human Universals* (New York: McGraw-Hill, 1991).

Buchanan, James and Gordon Tullock, *The Calculus of Consent: Logical Foundations of Constitutional Democracy* (Ann Arbor: University of Michigan Press, 1962).

— , Robert Tollison, and Gordon Tullock, *Toward a Theory of the Rent-seeking Society* (College Station: Texas A&M University Press, 1980).

Camerer, Colin, *Behavioral Game Theory: Experiments in Strategic Interaction* (Princeton, NJ: Princeton University Press, 2003).

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

Camerer, Colin F. and Ernst Fehr, "Measuring Social Norms and Preferences Using Experimental Games: A Guide for Social Scientists," in Joseph Henrich, Robert Boyd, Samuel Bowles, Colin F. Camerer, Ernst Fehr, and Herbert Gintis (eds.) *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Socieites* (Oxford: Oxford University Press, 2004) pp. 55–95.

Damasio, Antonio R., *Descartes' Error: Emotion, Reason, and the Human Brain* (New York: Avon Books, 1994).

Dawes, Robyn M., "Social Dilemmas," *Annual Review of Psychology* 31 (1980):169–193.

Downs, Anthony, *An Economic Theory of Democracy* (Boston: Harper & Row, 1957).

Eibl-Eibesfeldt, I., "Warfare, Man's Indoctrinability and Group Selection," *Journal of Comparative Ethnology* 60,3 (1982):177–198.

Ely, Jeffrey C. and Juuso Välimäki, "A Robust Folk Theorem for the Prisoner's Dilemma," *Journal of Economic Theory* 102 (2002):84–105.

Fehr, Ernst and Urs Fischbacher, "Third Party Punishment and Social Norms," *Evolution & Human Behavior* 25 (2004):63–87.

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

Fong, Christina M., "Social Preferences, Self-Interest, and the Demand for Redistribution," *Journal of Public Economics* 82,2 (2001):225–246.

— , Samuel Bowles, and Herbert Gintis, "Reciprocity and the Welfare State," in Herbert Gintis, Samuel Bowles, Robert Boyd, and Ernst Fehr (eds.) *Moral Sentiments and Material Interests: On the Foundations of Cooperation in Economic Life* (Cambridge: The MIT Press, 2005).

Frank, Robert H., "If *Homo Economicus* Could Choose His Own Utility Function, Would He Want One with a Conscience?," *American Economic Review* 77,4 (September 1987):593–604.

Friedman, Daniel and Nirvikar Singh, "Negative Reciprocity: The Coevolution of Memes and Genes," *Evolution and Human Behavior* 25,3 (May 2001):155–173.

Fudenberg, Drew and Eric Maskin, "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information," *Econometrica* 54,3 (May 1986):533–554.

— , David K. Levine, and Eric Maskin, "The Folk Theorem with Imperfect Public Information," *Econometrica* 62 (1994):997–1039.

Ghemawat, Pankaj, "Competitive Advantage and Internal Organization: Nucor Revisited," *Journal of Economic and Management Strategy* 3,4 (winter 1995):685–717.

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

— , "The Hitchhiker's Guide to Altruism: Genes, Culture, and the Internalization of Norms," *Journal of Theoretical Biology* 220,4 (2003):407–418.

— , "Modeling Cooperation among Self-Interested Agents: A Critique," *Journal of Socio-Economics* 33,6 (2004):697–717.

— , Samuel Bowles, Robert Boyd, and Ernst Fehr, *Moral Sentiments and Material Interests: On the Foundations of Cooperation in Economic Life* (Cambridge: The MIT Press, 2005).

Greif, Avner, "Cultural Beliefs and the Organization of Society: An Historical and Theoretical Reflection on Collectivist and Individualist Societies," *Journal of Political Economy* 102,5 (October 1994):912–950.

— , Paul Milgrom, and Barry R. Weingast, "Coordination, Commitment, and Enforcement: The Case of the Merchant Guild," *Journal of Political Economy* 104,4 (August 1994):745–776.

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

Güth, Werner, R. Schmittberger, and B. Schwarze, "An Experimental Analysis of Ultimatum Bargaining," *Journal of Economic Behavior and Organization* 3 (May 1982):367–388.

Hansen, Daniel G., "Individual Responses to a Group Incentive," *Industrial and Labor Relations Review* 51,1 (October 1997):37–49.

Harding, Susan, "Street Shouting and Shunning,: Conflict between Women in a Spanish village," *Frontiers* III,3 (1978):14–18.

Henrich, Joe, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, and Herbert Gintis, *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-scale Societies* (Oxford: Oxford University Press, 2004).

Hirshleifer, Jack, "Economics from a Biological Viewpoint," in Jay B. Barney and William G. Ouchi (eds.) *Organizational Economics* (San Francisco: Jossey-Bass, 1987) pp. 319–371.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler, "Fairness as a Constraint on Profit Seeking: Entitlements in the Market," *American Economic Review* 76,4 (September 1986):728–741.

Kandori, Michihiro, "Introduction to Repeated Games with Private Monitoring," *Journal of Economic Theory* 102 (2002):1–15.

Knez, Marc and Duncan Simester, "Firm-Wide Incentives and Mutual Monitoring at Continental Airlines," *Journal of Labor Economics* 19,4 (2001):743–772.

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.

Luttmer, Erzo F. P., "Group Loyalty and the Taste for Redistribution," *Journal of Political Economy* 109,3 (June 2001):500–528.

Mahdi, Niloufer Qasim, "Pukhtunwali: Ostracism and Honor Among the Pathan Hill Tribes," *Ethology and Sociobiology* 7 (1986):295–304.

Matsushima, H., "The Folk Theorem with Private Monitoring and Uniform Sustainability," 2000. CIRJE Discussion Paper F-84, University of Tokyo.

Moene, Karl Ove and Michael Wallerstein, "Inequality, Social Insurance and Redistribution," *American Political Science Review* 95,4 (2002):859–874.

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

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

Piccione, Michele, "The Repeated Prisoner's Dilemma with Imperfect Private Monitoring," *Journal of Economic Theory* 102 (2002):70–83.

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.

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

Sekiguchi, Tadashi, "Efficiency in Repeated Prisoner's Dilemma with Private Monitoring," *Journal of Economic Theory* 76 (1997):345–361.

Simon, Herbert, "A Mechanism for Social Selection and Successful Altruism," *Science* 250 (1990):1665–1668.

Stern, Jessica, *Terror in the Name of God* (New York: Harper Collins, 2003).

Taylor, Michael, *Anarchy and Cooperation* (London: John Wiley and Sons, 1976).

Trivers, Robert L., "The Evolution of Reciprocal Altruism," *Quarterly Review of Biology* 46 (1971):35–57.

Varshney, Ashutosh, "Nationality, Ethnic Conflict, and Rationality," *www.apsanet.org* 1,1 (2003):85–99.

Wiessner, Polly, "Norm Enforcement Among the Ju/'hoansi Bushmen: A Case of Strong Reciprocity?," *Human Nature* 16,2 (June 2005):115–145.

Wittman, Donald, "Why Democracies Produce Efficient Results," *Journal of Political Economy* 97,6 (1989):1395–1424.

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

Young, Peyton and Mary Burke, "Competition and Custom in Economic Contracts: A Case Study of Illinois Agriculture," *American Economic Review* 91,3 (2001):559–573.

TABLE 1. Seven experimental games useful for measuring social preferences

Game	Definition of the game	Real life example	Predictions with rational and selfish players	Experimenta regularities, references	Interpretation
Prisoner's Dilemma Game	Two players, each of whom can either cooperate or defect. Payoffs are as follows: Cooperate: Cooperate H,H; Defect S,T Defect: T,S; L,L $H > L$, $T > H$, $L > S$	Production of negative externalities (pollution, loud noise), exchange without binding contracts, status competition.	Defect.	50% choose to cooperate. Communication increases frequency of cooperation. Dawes (1980).[a]	Reciprocate expected cooperation.
Public Goods Game	n players simultaneously decide about their contribution g_i. $(0 \leq g_i \leq y)$ where y is players' endowment; each player i earns $\pi_i = y - g_i + mG$ where G is the sum of all contributions and $m < 1 < mn$.	Team compensation, cooperative production in simple societies, overuse of common resources (e.g. water, fishing grounds).	Each player contributes nothing, that is, $g_i = 0$.	Players contribute 50% of y in the one-shot game. Contributions unravel over time. Majority chooses $g_i = 0$ in final period. Communication strongly increases cooperation. Individual punishment opportunities greatly increase contributions. Ledyard (1995).[a]	Reciprocate expected cooperation.

TABLE 1. *Continued*

Game	Definition of the game	Real life example	Predictions with rational and selfish players	Experimenta regularities, references	Interpretation
Ultimatum Game	Division of a fixed sum of money S between a proposer and a responder. proposer offers x. If responder rejects x both earn zero, if x is accepted the proposer earns $S-x$ and the responder earns x.	Monopoly pricing of a perishable good; '11th hour' settlement offers before a time deadline.	Offer $x=\varepsilon$ where ε is the smallest money unit. Any $x>0$ is accepted.	Most offers are between 0.3 and 0.5S. $x<0.2S$ rejected half the time. Competition among proposers has a strong x-increasing effect; competition among responders strongly decreases x. Güth, Schmitterbert and Schwartze (1982),[b] Camerer (2003).[a]	Responders punish unfair offers; negative reciprocity.
Dictator Game	Like the UG but the responder cannot reject, that is, the 'proposer' dictates $(S-x, x)$.	Charitable sharing of a windfall gain (lottery winners giving anonymously to strangers).	No sharing, that is, $x=0$.	On average 'proposers' allocate $x=0.2S$. Strong variations across experiments and across individuals. Kahneman, Knetsch, and Thaler (1986),[b] Camerer (2003).[a]	Pure altruism.

Trust Game	Investor has endowment S and makes a transfer y between 0 and S to the trustee. Trustee receives $3y$ and can send back any x between 0 and $3y$. Investor earns $S-y+x$. Trustee earns $3y-x$.	Sequential exchange without binding contracts (buying from sellers on e-Bay).	Trustee repays nothing: $x=0$. Investor invests nothing: $y=0$.	On average $y=0.5S$ and trustees repay slightly less than $0.5S$. x is increasing in y. Berg Dickhaut, and McCabe (1995),[b] Camerer (2003).[a]	Trustees show positive reciprocity.
Gift Exchange Game	'Employer' offers a wage w to the 'Worker' and announces a desired effort level $\hat{e}$. If worker rejects $(w, \hat{e})$ both earn nothing. If Worker accepts, he can choose *any* e between 1 and 10. Then Employer earns $10e-w$ and Worker earn $w-c(e)$. $c(e)$ is the effort cost which is strictly increasing in e.	Noncontractibility or nonenforceability of the performance (effort, quality of goods) of workers or sellers.	Worker chooses $e=1$. Employer pays the minimum wage.	Effort increases with the wage w. Employers pay wages that are far above the minimum. Workers accept offers with low wages but respond with $e=1$. In contrast to the UG competition among workers (i.e. responders) has no impact on wage offers. Fehr, Kirchsteiger, and Riedl (1993).[b]	Workers reciprocate generous wage offers. Employers appeal to workers' reciprocity by offering generous wages.
Third-Party Punishment Game	A and B play a DG. C observes how much of amount S is allocated to B. C can punish A but the punishment is also costly for C.	Social disapproval of unacceptable treatment of others (scolding neighbors).	A allocates nothing to B. C never punishes A.	Punishment of A is higher, the less A allocates to B. Fehr and Fischbacher (2001*a*).[b]	C sanctions violation of a sharing norm.

Note: [a] Denotes survey papers. [b] denotes papers that introduced the respective games.