Cooperation

Samuel Bowles and Herbert Gintis*

April 10, 2007

Abstract

We review game-theoretic models of cooperation with self-regarding agents, starting with retaliation against noncooperators and reputation-building in repeated dyadic interactions. We then study the Folk Theorem in large groups of self-regarding individuals with imperfect information. In contrast to the dyadic case with perfect information, we find that the level of cooperation deteriorates rapidly with larger group size and higher error rates. Moreover, there is no plausible account of how the dynamic, out-of-equilibrium, behavior of these models would support cooperative outcomes. We then analyze cooperation with other-regarding preferences, finding that a high level of cooperation can be attained in groups of substantial size and with modest informational requirements, and that conditions allowing the long term evolution of such social preferences are plausible.

1 Introduction

Cooperation is said to occur when two or more individuals engage in joint actions that result in mutual benefits. Examples include the mutually beneficial exchange of goods, the payment of taxes to finance public goods, team production, common pool resource management, collusion among firms, voting for income redistribution to others, participating in collective actions such as demonstrations, and adhering to socially beneficial norms.

A major goal of economic theory has been to explain how wide-scale cooperation among self-regarding individuals occurs in a decentralized setting. The first thrust

*To appear in L. Blume and S. Durlauf, eds., *The New Palgrave Dictionary of Economics* (MacMillan 2008). This research was supported by the Behavioral Sciences Program of the Santa Fe Institute, the Russell Sage Foundation, the MacArthur Foundation, and the University of Siena. Thanks to Margaret Alexander, Robert Boyd, Eric Maskin, Martin Nowak, E. Somanathan, Timothy Taylor, Della Ulibarri and Elisabeth Wood for their contributions. Affiliations: Bowles, Santa Fe Institute and University of Siena, Gintis: Santa Fe Institute and Central European University.

of this endeavor involved Walras' general equilibrium model, culminating in the celebrated 'invisible hand' theorem of Arrow and Debreu (Arrow and Debreu 1954, Debreu 1959, Arrow and Hahn 1971). But, the assumption that contracts could completely specify all relevant aspects of all exchanges and could be enforced at zero cost to the exchanging parties is not applicable to many important forms of cooperation. Indeed, such economic institutions as firms, financial institutions, and state agencies depend on incentive mechanisms involving strategic interaction in addition to explicit contracts (Blau 1964, Gintis 1976, Stiglitz 1987, Tirole 1988, Laffont 2000).

The second major thrust in explaining cooperation eschewed complete contracting and developed sophisticated repeated game-theoretic models of strategic interaction. These models are based on the insights of Shubik (1959), Taylor (1976), Axelrod and Hamilton (1981) and others that repetition of social interactions plus retaliation against defectors by withdrawal of cooperation may enforce cooperation among self-regarding individuals. A statement of this line of thinking, applied towards understanding the broad historical and anthropological sweep of human experience is the work of Ken Binmore (1993, 1998, 2005), For Binmore, a society's moral rules are instructions for behavior in conformity with one of the myriad of Nash equilibria of a repeated n-player social interaction. Because the interactions are repeated, and these rules form a Nash equilibrium, the self-regarding individuals who comprise the social order will conform to the moral rules.

We begin by reviewing models of repeated dyadic interaction in which cooperation may occur among players who initially cooperate and in the next round adopt the action of the other player in the previous round, called *tit-for-tat*. These models show that as long as the probability of game repetition is sufficiently great and individuals are sufficiently patient, a cooperative equilibrium can be sustained once it is implemented. This reasoning applies to a wide range of similar strategies. We then analyze *reputation maintenance* models of dyadic interaction, which are relevant when individuals interact with many different individuals, and hence the number of periods before a repeat encounter with any given individual may be too great to support the tit-for-tat strategy.

We then turn to models of cooperation in larger groups, arguably the most relevant case, given the scale on which cooperation frequently takes place. The *Folk Theorem* (Fudenberg and Maskin 1986) shows that in groups of any size, cooperation can be maintained assuming the players are sufficiently future-oriented and termination of the interaction is sufficiently unlikely. We will see, however, that these models do not successfully extend the intuitions of the dyadic models to many-person interactions. The reason is that the level of cooperation that may be supported in this way deteriorates as group size increases and the probability of either behavioral or perceptual error rises, and because the theory lacks a plausible account how individuals would discover and coordinate on the complicated strategies necessary for cooperation to be sustained in these models. This difficulty bids us investigate how other-regarding preferences, *strong reciprocity* in particular, may sustain a high level of cooperation, even with substantial errors and in large groups.

2 Repetition Allows Cooperation in Groups of Size Two

Consider a pair of individuals who play the following *stage game* repeatedly: each can *cooperate* (i.e., help the other) at a cost $c > 0$ to himself, providing a benefit to the other of $b > c$. Alternatively, each player can *defect*, incurring no cost and providing no benefit. Clearly, both would gain by cooperating in the stage game, each receiving a net gain of $b - c > 0$. However, the structure of the game is that of a *prisoner's dilemma*, in which a self-regarding player earns higher payoff by defecting, no matter what his partner does.

The behavior whereby each individual provides aid as long as this aid has been reciprocated by the other in the previous encounter, is called *tit-for-tat*. Although termed 'reciprocal altruism' by biologists, this behavior is self-regarding, because each individual's decisions depend only on the expected net benefit the individual enjoys from the long-term relationship.

Assuming that after each round of play the interaction will be continued with probability δ, and assuming players have discount factor d (so $d = 1/(1+r)$, where r is the rate of time preference), then provided

$$\delta d b > c, \tag{1}$$

each individual paired with a tit-for-tat player does better by cooperating (that is, playing tit-for-tat) rather than by defecting. Thus tit-for-tat is a best response to itself. To see this, let $\mathbf{v}$ be the present value of cooperating when paired with a tit-for-tat player. Then

$$\mathbf{v} = b - c + \delta d \mathbf{v}, \tag{2}$$

which gives

$$\mathbf{v} = \frac{b - c}{1 - \delta d}. \tag{3}$$

The present value of defecting forever on a tit-for-tat playing partner is b (the single period gain of b being followed by zero gains in every subsequent period as a result of the tit-for-tat player's defection), so playing tit-for-tat is a best response to itself if and only if $(b - c)(1 - \delta d) > b$, which reduces to (1). Under these conditions unconditional defect is also a best response to itself, so either cooperation or defection can be sustained.

But suppose that instead of defection forever, the alternative to tit-for-tat is for a player to defect for a certain number of rounds, before returning to cooperation on round $k > 0$. The payoff to this strategy against tit-for-tat is $b - (\delta d)^k c + (\delta d)^{k+1}\mathbf{v}$. This payoff must not be greater than $\mathbf{v}$ if tit-for-tat is to be a best response to itself. It is an easy exercise in algebra to show that the inequality

$$\mathbf{v} \geq b - (\delta d)^k c + (\delta d)^{k+1}\mathbf{v}$$

simplifies to (1), no matter what the value of k. A similar argument shows that when (1) holds, defecting forever (i.e., $k = \infty$) does not have a higher payoff than cooperating.

3 Cooperation Through Reputation Maintenance

Tit-for-tat takes the form of frequent repetition of the prisoner's dilemma stage game inducing a pair of self-regarding individuals to cooperate. In a sizeable group, an individual may interact frequently with a large number of partners, but infrequently with any single one, say on the average of once every k periods. Players then discount future gains so that a payoff of $\mathbf{v}$ in k periods from now is worth $d^k\mathbf{v}$ now. Then, an argument parallel to that of the previous section shows that cooperating is a best response if and only if

$$\frac{b-c}{1-\delta d^k} > b$$

which reduces to

$$\delta d^k b > c. \tag{4}$$

Note that this is the same equation as (1) except that the effective discount factor falls from d to d^k. For sufficiently large k, it will not pay to cooperate. Therefore, the conditions for tit-for-tat reciprocity will not obtain.

But, cooperation may be sustained in this situation if each individual keeps a mental model of exactly which group members cooperated in the previous period and who did not. In this case, players may cooperate in order to cultivate a *reputation for cooperation*. When individuals tend to cooperate with others who have a reputation for cooperation, a process called *indirect reciprocity* can sustain cooperation. Let us say that an individual who cooperated in the previous period *in good standing*, and specify that the only way an individual can fall into *bad standing* is by defecting on a partner who is in good standing. Note that an individual can always defect when his partner is in bad standing without losing his good standing status. In this more general setting the tit-for-tat strategy is replaced by the following *standing strategy*: cooperate if and only if your current partner is in good standing, except that if you accidentally defected the previous period, cooperate this period unconditionally, thereby restoring your status as a member in good standing. This *standing model* is due to Sugden (1986),

Panchanathan and Boyd (2004) have proposed an ingenious deployment of indirect reciprocity, assuming that there is an on-going dyadic helping game in society based on the indirect reciprocity information and incentive structure, and there is also an n-player public goods game, played relatively infrequently by the same individuals. In the dyadic helping game, two individuals are paired and each member of the pair may confer a benefit b upon his partner at a cost c to himself, an individual remaining in good standing so long as he does not defect on a partner who is in good standing. This random pairing is repeated with probability δ and with discount factor d. In the public goods game, an individual produces a benefit b_g that is shared equally by all the other members, at a cost c_g to himself. The two games are linked by defectors in the public goods game being considered in bad standing at the start of the helping game that directly follows. Then, cooperation can be sustained in both the public goods game and in the dyadic helping game so long as

$$c_g \leq \frac{b(1-\epsilon)-c}{1-\delta d}, \tag{5}$$

where ϵ is the rate at which cooperators unintentionally fail to produce the benefit. Parameters favoring this solution are that the cost c_g of cooperating in the public goods game be low, the factor δd is close to unity, and the net benefit $b(1-\epsilon)-c$ of cooperating in the reputation-building reciprocity game be large.

The major weakness of the standing model is its demanding informational requirements. Each individual must know the current standing of each member of the group, the identity of each member's current partner, and whether each individual cooperated or defected against his current partner. Since dyadic interactions are generally private, and hence are unlikely to be observed by more than a small number of others, errors in determining the standing of individuals may be frequent. This contrasts sharply with the repeated game models of the previous section, which require only that an individual know how many of his current partners defected in the previous period. Especially serious is that warranted non-cooperation (because in one's own mental accounting one's partner is in bad standing) may be perceived to be unwarranted defection by some third parties but not by others. This will occur with high frequency if information partially private rather than public (not everyone has the same information). It has been proposed that gossip and other forms of communication can transform private into public information, but how this might occur among self-regarding individuals has not been (and probably cannot be) shown, because in any practical setting individuals may benefit by reporting dishonestly on what they have observed, and self-regarding individuals do not care about the harm to others induced by false information. Under such conditions, disagreements among individuals about who ought to be punished can reach extremely high levels, with the unraveling of cooperation as a result.

In response to this weakness of the standing model, Nowak and Sigmund (1998) developed an indirect reciprocity model which they term *image scoring*. Players in the image-scoring need not know the standing of recipients of aid, so the informational requirements of indirect reciprocity are considerably reduced. Nowak and Sigmund show that the strategy of cooperating with others who have cooperated in the past, *independent of the reputation of the cooperator's partner*, is stable against invasion by defectors, and weakly stable against invasion by unconditional cooperators once defectors are eliminated from the population. Leimar and Hammerstein 2001, Panchanathan and Boyd 2003, and Brandt and Sigmund 2004, 2005, explore the applicability of image scoring.

4 Cooperation in Large Groups of Self-Regarding Individuals

Repeated game theory has extended the above two-player results to a general n-player stage game, the so-called *public goods game*. In this game each player cooperates at cost $c > 0$, contributing an amount $b > c$ that is shared equally among the other $n - 1$ players. We define the *feasible payoff set* as the set of possible payoffs to the various players, assuming each cooperates with a certain probability, and each player does at least as well as the payoffs obtaining under mutual defection. The set of feasible payoffs for a two-player public goods game is given in Figure 1 by the four-sided figure ABCD. For the n-player game, the figure ABCD is replaced by a similar n-dimensional polytope.

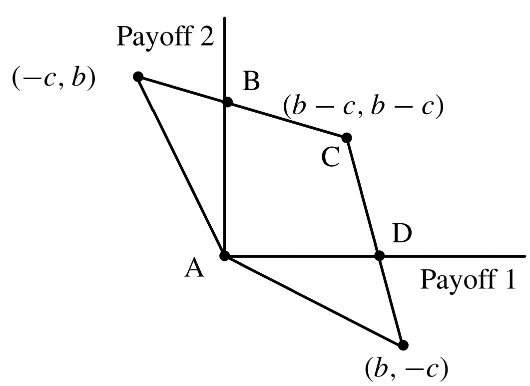

Figure 1: Two-player Public Goods Game

Repeated game models have demonstrated the so-called *Folk Theorem*, which asserts that *any distribution of payoffs to the n players that lies in the feasible payoff* *set can be supported by an equilibrium in the repeated public goods game, provided the discount factor times the probability of continuation, δd, is sufficiently close to unity*. The equilibrium concept employed is a refinement of subgame perfect equilibrium. Significant contributions to this literature include Fudenberg and Maskin (1986), assuming perfect information, Fudenberg, Levine and Maskin (1994), assuming imperfect information, so that cooperation is sometimes inaccurately reported as defection, and Sekiguchi (1997), Piccione (2002), Ely and Välimäki (2002), Bhaskar and Obara (2002), and Mailath and Morris (2006), who assume that different players receive different, possibly inaccurate, information concerning the behavior of the other players.

The Folk Theorem is an *existence theorem* affirming that any outcome that is a Pareto improvement over universal defection may be supported by a Nash equilibrium, including point C (full cooperation) in the figure and outcomes barely superior to A (universal defection). The theorem is silent on which of these vast number of equilibria is more likely to be observed or how they might be attained. When these issues are addressed two problems are immediately apparent: first, equilibria in the public goods game supported in this manner exhibit very little cooperation if large numbers of individuals are involved or errors in execution and perception are large, and second, the equilibria are not robust because they require some mechanism allowing coordination on highly complex strategies. While such a mechanism could be provided by centralized authority, decentralized mechanisms, as we will see, are not sustainable in a plausible dynamic.

5 The Dynamics of Cooperation

The first difficulty, the inability to support high levels of cooperation in large groups or with significant behavioral or perceptual noise stems from the fact that the only way players may punish defectors is *to withdraw their own cooperation*. In the two-person case, defectors are thus targeted for punishment. But for large n, withdrawal of cooperation to punish a single defector punishes all group members equally, most of whom, in the neighborhood of a cooperative equilibrium, will be cooperators. Moreover, in large groups, the rate at which erroneous signals are propagated will generally increase with group size, and the larger the group, the larger the fraction of time group members will spend punishing (miscreants and fellow cooperators alike). For instance, suppose the rate at which cooperators accidentally fail to produce b, and hence signal defection, is 5%. Then, in a group of size two, a perceived defection will occur in about 10% of all periods, while in a group of size 20, at least one perceived defection will occur in about 64% of all periods.

As a result of these difficulties, the Folk Theorem assertion that we can approximate the per-period expected payoff as close to the efficient level (point C in Figure 1) as desired as long as the discount factor δ is sufficiently close to unity is of little practical relevance. The reason is that as $\delta \to 1$, the current payoff approximates zero, and the expected payoff is deferred to future periods at very little cost, since future returns are discounted at a very low rate. Indeed, holding constant the discount factor δ, the efficiency of cooperation in the Fudenberg, Levine, and Maskin model declines at an exponential rate with increasing group size (Bowles and Gintis 2007, Ch. 13). Moreover, in an agent-based simulation of the public goods with punishment model, assuming a benefit/cost ratio of $b/c = 2$ (i.e., contributing to the public good costs half of the benefit conferred on members of the group) and a discount factor times probability of repetition of $d\delta = 0.96$, even for an error rate as low as $\epsilon = 0.04$, there is fewer than half of the members contribute to the public good in groups of size $n = 4$, and less that 20% contribute in groups of size $n = 6$ (Bowles and Gintis 2007, Ch. 5).

The second limitation of the Folk Theorem analysis is that it has not been shown (and probably cannot be shown) that the equilibria supporting cooperation are dynamically robust; i.e., asymptotically stable with a large basin of attraction in the relevant dynamic. Equilibria for which this is not the case will seldom be observed because they are unlikely to be attained and if attained unlikely to persist for long.

The Nash equilibrium concept applies when each individual expects all others to play their parts in the equilibrium. But, when there are multiple equilibria, as in the case of the Folk Theorem, where there are many possible patterns of response to given pattern of defection, each imposing distinct costs and requiring distinct, possibly stochastic, behaviors on the part of players, there is no single set of beliefs and expectations that group members can settle upon to coordinate their actions (Aumann and Brandenburger 1995).

While game theory does not provide an analysis of how beliefs and expectations are aligned in a manner allowing cooperation to occur, sociologists (Durkheim 1967[1902], Parsons and Shils 1951) and anthropologists (Benedict 1934, Boyd and Richerson 1985, Brown 1991) have found that virtually every society has such processes, and that they are key to understanding strategic interaction. Borrowing a page from sociological theory, we posit that groups may have *focal rules* that are common knowledge among group members. Focal rules could suggest which of a countless number of strategies that could constitute a Nash equilibrium should all individuals adopt them, thereby providing the coordination necessary to support cooperation. These focal rules do not ensure equilibrium, because error, mutation, migration, and other dynamical forces ensure that on average not all individuals conform to the focal rules of the groups to which they belong. Moreover, a group's focal rules are themselves subject to dynamical forces, those producing better outcomes for their members displacing less effective focal rules.

In the case of the repeated public goods game, which is the appropriate model for many forms of large-scale cooperation, Gintis (2007) shows that focal rules capable of supporting the kinds of cooperative equilibria identified by the Folk Theorem are not evolutionarily stable, meaning that groups whose focal rules support highly cooperative equilibria do worse than groups with less stringent focal rules, and as a result the focal rules necessary for cooperation are eventually eliminated.

The mechanism behind this result can be easily explained. Suppose a large population consists of many smaller groups playing n-person public goods games, with considerable migration across groups, and with the focal rules of successful groups being copied by less successful groups. To maintain a high level of cooperation in a group, focal rules should foster punishing defectors by withdrawing cooperation. However, such punishment is both costly and provides an external benefit to other groups by reducing the frequency of defection-prone individuals who might migrate elsewhere. Hence, groups that "free ride" by not punishing defectors harshly will support higher payoffs for its members than groups that punish assiduously. Such groups will then be copied by other groups, leading to a secular decline in the frequency of punishment suggested by focal rules in all groups. Thus, suppose that the groups in question were competitive firms whose profits depend on the degree of cooperation among firm members. If all adopted a zero-tolerance rule (all would defect if even a single defection was perceived), then a firm adopting a rule that tolerated a single defection would sustain higher profits and replace the zero-tolerance firms. But this firm would in turn be replaced by a firm with a tolerate two defections rule.

These two problems—the inability to support efficient levels of cooperation in large groups with noisy information, and dynamic instability—have been shown for the case where information is public. Private information, in general the more relevant case, considerably exacerbates these problems.

6 Cooperation with Other-regarding Individuals

The models reviewed thus far have assumed that individuals are entirely self-regarding. But cooperation in sizable groups is possible if there exist other-regarding individuals in the form of *strong reciprocators*, who cooperate with one another, and punish defectors, even if they sustain net costs. Strong reciprocators are altruistic in the standard sense that they confer benefits on other members of their group (in this case, because their altruistic punishment of defectors sustains cooperation) but would increase their own payoffs by adopting self-regarding behaviors. A model with *social preferences* of this type can explain large-scale decentralized cooperation with noisy information as long as the information structure is such that defectors expect a level of punishment greater than costs of cooperating.

Cooperation is not a puzzle if a sufficient number of individuals with social preferences are involved. The puzzle that arises is how such altruistic behavior could have become common, given that bearing costs to support the benefits of others reduces payoffs, and both cultural and genetic updating of behaviors is likely to favor traits with higher payoffs. This evolutionary puzzle applies to strong reciprocity. Since punishment is costly to the individual, and an individual could escape punishment by cooperating, while avoiding the costs of punishment by not punishing, we are obliged to exhibit a mechanism whereby strong reciprocators could proliferate when rare and be sustained in equilibrium, despite their altruistic behavior.

This is carried out in Sethi and Somanathan (2001), Gintis (2000), Boyd, Gintis, Bowles and Richerson (2003), Gintis (2003), and Bowles and Gintis (2004). The evolutionary viability of other types of altruistic cooperation is demonstrated in Bowles, Choi and Hopfensitz (2003), Boyd et al. (2003), Bergstrom (1995), and Salomonsson and Weibull (2006). The critical condition allowing the evolution of strong reciprocity and other forms of altruistic social preferences is that individuals with social preferences are more likely than random to interact with others with social preferences. Positive assortment arises in these models due to deliberate exclusion of those who have defected in the past (by ostracism, for example), random differences in the composition of groups (due to small group size and limited between-group mobility), limited dispersion of close kin who share common genetic and cultural inheritance, and processes of social learning such as conformism or group level socialization contributing to homogeneity within groups. As in the repeated game models, smaller groups favor cooperation, but in this case for a different reason: positive assortment tends to decline with group size. But the group sizes that sustain the altruistic preferences that support cooperative outcomes in these models are at least an order of magnitude larger than those indicated for the repeated game models studied above.

In sum, we think that other-regarding preferences provide a compelling account of many forms of human cooperation that are not well explained by repeated game models with self-regarding preferences. Moreover, a number of studies have shown that strong reciprocity and other social preferences are a common human behavior (Fehr and Gächter 2000, Henrich, Boyd, Bowles, Camerer, Fehr and Gintis 2005) and could have emerged and been sustained in a gene-culture coevolutionary dynamic under conditions experienced by ancestral humans (Bowles 2006). The above models also show that strong reciprocity and other social preferences that support cooperation can evolve and persist even when there are many self-regarding players, where group sizes are substantial, and when behavioral or perception errors are significant.

7 Conclusion: Economics and the Missing Choreographer

The shortcomings of the economic theory of cooperation based on repeated games strikingly replicate those of economists' other main contribution to the study of decentralized cooperation, general equilibrium theory. Both prove the existence of equilibria with socially desirable properties, while leaving the question of how such equilibria are achieved as an afterthought, thereby exhibiting a curious lack of attention to dynamics and out-of equilibrium behavior. Both purport to model decentralized interactions but on close inspection require a level of coordination that is not explained, but rather posited as a *deus ex machina*. To ensure that only equilibrium trades are executed, general equilibrium theory resorts to a fictive "auctioneer." No counterpart to the auctioneer has been made explicit in the repeated game approach cooperation. Highly choreographed coordination on complex strategies capable of deterring defection are supposed to materialize quite without the need for a choreographer.

Humans are unique among living organisms in the degree and range of cooperation among large numbers of substantially unrelated individuals. The global division of labor and exchange, the modern democratic welfare state, and contemporary warfare alike evidence our distinctiveness. These forms of cooperation emerged historically and are today sustained as a result of the interplay of self-regarding and social preferences operating under the influence of group level institutions of governance and socialization that favor cooperators, in part by protecting them from exploitation by defectors.

The norms and institutions that have accomplished this evolved over millennia through trial and error. Consider how real-world institutions addressed two of the shoals on which the economic models foundered. First, the private nature of information, as we have seen, makes it virtually impossible to coordinate the targeted punishment of miscreants. Converting private information about transgressions to public information that can provide the basis of punishment often involves civil or criminal trials, elaborate processes that rely on commonly agreed upon rules of evidence and ethical norms of appropriate behavior. Even these complex institutions frequently fail to transform the private protestations of innocence and guilt into common knowledge. Second, as in the standing models with private information, cooperation often unravels when the withdrawal of cooperation by the civic-minded intending to punish a defector is interpreted by others as a violation of a cooperative norm, inviting further defections. In all successful modern societies, this problem was eventually addressed by the creation of a corps of specialists entrusted with carrying out the more severe of society's punishments, whose uniforms conveyed the civic purpose of the punishments they meted out, and whose professional norms, it was hoped, would ensure that the power to punish was not used for personal gain.

Like court proceedings, this institution works imperfectly.

It is hardly surprising then that economists have encountered difficulty in devising simple models of how large numbers of self-regarding individuals might sustain cooperation in a truly decentralized setting. Modeling this complex process is a major challenge of contemporary science. Economic theory, favoring parsimony over realism, has instead sought to explain cooperation without reference to other-regarding preferences and with minimalist or fictive descriptions of social institutions.

References

Arrow, Kenneth J. and Frank Hahn, *General Competitive Analysis* (San Francisco: Holden-Day, 1971).

— and Gerard Debreu, "Existence of an Equilibrium for a Competitive Economy," *Econometrica* 22,3 (1954):265–290.

Aumann, Robert J. and Adam Brandenburger, "Epistemic Conditions for Nash Equilibrium," *Econometrica* 65,5 (September 1995):1161–80.

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

Benedict, Ruth, *Patterns of Culture* (Boston: Houghton Mifflin, 1934).

Bergstrom, Theodore C., "On the Evolution of Altruistic Ethical Rules for Siblings," *American Economic Review* 85,1 (March 1995):58–81.

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

Binmore, Ken, *Game Theory and the Social Contract: Playing Fair* (Cambridge, MA: MIT Press, 1993).

— , *Game Theory and the Social Contract: Just Playing* (Cambridge, MA: MIT Press, 1998).

Binmore, Kenneth G., *Natural Justice* (Oxford: Oxford University Press, 2005).

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

Bowles, Samuel, "Group Competition, Reproductive Leveling, and the Evolution of Human Altruism," *Science* 314 (8 December 2006):1669–72.

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

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

— , 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, *Culture and the Evolutionary Process* (Chicago: University of Chicago Press, 1985).

— , 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.

Brandt, Hannalore and Karl Sigmund, “The Logic of Reprobation: Assessment and Action Rules for Indirect Reciprocation,” *Journal of Theoretical Biology* 231 (2004):475–486.

— and — , “Indirect Reciprocity, Image Scoring, and Moral Hazard,” *Proceeding of the National Academy of Sciences* 102,7 (February 15 2005):2666–2670.

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

Debreu, Gérard, *Theory of Value* (New York: Wiley, 1959).

Durkheim, Emile, *De La Division du Travail Social* (Paris: Presses Universitaires de France, 1967[1902]).

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

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.

Gintis, Herbert, “The Nature of the Labor Exchange and the Theory of Capitalist Production,” *Review of Radical Political Economics* 8,2 (Summer 1976):36–54.

— , “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 with Self-regarding Agents,” 2007. Santa Fe Institute.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, and Herbert Gintis, “Economic Man’ in Cross-Cultural Perspective: Behavioral Experiments in 15 small-scale societies,” *Behavioral and Brain Sciences (*2005).

Laffont, Jean Jacques, *Incentives and Political Economy* (Oxford: Oxford University Press, 2000).

Leimar, O. and P. Hammerstein, “Evolution of Cooperation through Indirect Reciprocity,” *Proc. Royal. Soc. Lond. B* 268 (2001):745–753.

Mailath, George J. and Stephen Morris, “Coordination Failure in Repeated Games with Almost-public Monitoring,” *Theoretical Economics* 1 (2006):311–340.

Nowak, Martin A. and Karl Sigmund, “Evolution of Indirect Reciprocity by Image Scoring,” *Nature* 393 (1998):573–577.

Panchanathan, Karthik and Robert Boyd, “A Tale of Two Defectors: The Importance of Standing for Evolution of Indirect Reciprocity,” *Journal of Theoretical Biology* 224 (2003):115–126.

— and — , “Indirect Reciprocity can Stabilize Cooperation Without the Second-order Free Rider Problem,” *Nature* 432 (2004):499–502.

Parsons, Talcott and Edward Shils, *Toward a General Theory of Action* (Cambridge, MA: Harvard University Press, 1951).

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

Salomonsson, Marcus and Jörgen Weibull, “Natural Selection and Social Preferences,” *Journal of Theoretical Biology* 239 (2006):79–92.

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

Sethi, Rajiv and E. Somanathan, “Preference Evolution and Reciprocity,” *Journal of Economic Theory* 97 (2001):273–297.

Shubik, Martin, *Strategy and Market Structure: Competition, Oligopoly, and the Theory of Games* (New York: Wiley, 1959).

Stiglitz, Joseph, “The Causes and Consequences of the Dependence of Quality on Price,” *Journal of Economic Literature* 25 (March 1987):1–48.

Sugden, Robert, *The Economics of Rights, Co-operation and Welfare* (Oxford: Basil Blackwell, 1986).

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

Tirole, Jean, *The Theory of Industrial Organization* (Cambridge, MA: MIT Press, 1988).

c\Papers\palgrave-blume-durlauf\Palgrave.tex April 10, 2007