Altruistic Punishment in Politics and Life Sciences: Climbing the Same Mountain in Theory and Practice

Oleg Smirnov

As reflected in theory, laboratory evidence, and field studies, altruistic punishment of defectors promotes cooperation. Costly self-enforcement of socially optimal behavior has a number of independent links in political science, economics, psychology, sociology, computer science, and biology. This paper integrates the study of sanctions-based provision of public goods in the social sciences with the research on evolutionary adaptedness of altruistic punishment in the life sciences. Altruistic punishment appears to be (1) economically rational, (2) evolutionarily robust as an individual propensity and as a cultural norm, (3) normatively more appealing than tit-for-tat, which is a reciprocal punishment by defection, and (4) socially common. The theoretical and empirical importance of altruistic punishment has immediate policy implications. Examination of commons around the world suggests that privatization and centralized coercion are not the only solutions to the tragedy of the commons. A viable policy alternative is to facilitate the evolution of the commons as a common-property regime with its own norms and a certain degree of independence.

Altruistic behavior provides benefits to others while incurring a cost for the acting individual.[1] The existence of such behavior has been a puzzle for scholars across the social and life sciences. To social scientists, choosing A over B is irrational if B provides higher individual utility as does non-altruism. To life scientists, it is no less puzzling how altruism could have survived in the evolutionary process. Formal models of altruistic and selfish behavior, most notably the Prisoners' Dilemma game and public good games, have further increased interest in the paradox and led to an unprecedented surge of interdisciplinary research.[2] Early classical models predicted that rational behavior would lead to mutual defection, overexploitation of resources, and, ultimately, decline into a Hobbesian world when the life of a man is "solitary, poor, nasty, brutish, and short."[3] Strict dominance of selfish behavior made Hardin famously declare that there is no technical solution to the tragedy of the commons.[4]

Theoretical predictions of selfish behavior, however, are not supported by experimental research on public good games, which provides ambiguous results. In laboratory settings, subjects choose both cooperation and defection.[5] Similarly, field research indicates that successful management of common-pool resources is not rare even in the absence of Leviathan.[6] These facts add fuel to the theoretical examination of the puzzle of human cooperation. In the social sciences, the emphasis has been on game theoretic and computational models of the iterated n-person Prisoners' Dilemma game.[7] In this context, cooperation can be a product of prospectively rational individual choice if future interactions are contingent upon the history of past plays and the future discounting is not extreme. In the life sciences, scholars have been more concerned with conditions under which cooperation can be adaptive in the evolutionary process, or "retrospectively rational." In this case, altruism is viewed as a genetic trait that individuals inherit according to their fitness differential.[8] Nowadays some of the research also takes place across disciplines. Notable examples are the models of evolutionary game theory, models of the cultural evolution, and a variety of agent-based models.[9] Nowak presents a comprehensive overview of interdisciplinary models of evolution of cooperation.[10]

These interdisciplinary advances, however, have had a limited effect on actual policies for the management of common property resources. In order to avoid the tragedy of the commons, two standard policy prescriptions are adopted: non-altruism privatization of the common-pool resources, and centralized coercion by the government.[11] Both policies presume that cooperation is impossible. Any possibility of decentralized self-enforcement of

Oleg Smirnov is Assistant Professor of Political Science, Stony Brook University (Oleg.Smirnov@sunysb.edu). *He would like to thank Terry Anderson, Daniel Benjamin, James Fowler, Tim Johnson, John Orbell, Tony Smith, Wally Thurman, and anonymous referees for helpful comments. This research was supported by the Property and Environment Research Center (PERC), Bozeman, MT.*

DOI: 10.1017/S1537592707071502

cooperation is quickly dismissed as not viable; individual costs of monitoring and enforcement can be very high while the benefits are divided among members of the commons. Policing, therefore, is also a public good, and the problem of collective action is merely shifted one step higher.[12] As a result, the study of altruism has only focused on the cooperation strategy and, until recently, has ignored altruistic punishment as a redundant problem.

This article introduces the broad political science audience to the interdisciplinary research on altruistic punishment, which should aid our discipline in normative and positive analysis of the collective action problems. This work integrates the study of decentralized self-enforcement of cooperative norms as developed in the social sciences with research on evolutionary adaptedness of altruistic punishment as developed in the life sciences. The theoretical discussion of altruistic punishment is based upon rational choice and evolutionary game theoretic models. The parameters of such models, their underlying assumptions, and the conditions necessary for the maintenance of cooperation in them are further discussed through the prism of field studies in political science and laboratory experiments in economics and evolutionary psychology. On the basis of such analysis, the paper concludes with a normative discussion and suggestions of practical policy implications. Methodologically, the paper makes an attempt to show certain advantages of academic collaboration between political scientists and scholars representing other disciplines, from economics and computer science to biology and evolutionary psychology.

Definition of Altruistic Punishment

Altruistic punishment (AP) is a strategy according to which a cooperative individual punishes those who defect at a personal cost to oneself. Typically, but not necessarily, the cost of punishment is greater for the recipient (defector) than for the sender (cooperator who punishes). Under a common-property regime, altruistic punishers consume available resources maximizing social utility, and punish at a personal cost those individuals who overexploit common resources. Examples of such costly punishment include exiting mutually beneficial relationships, gossip, quarrel, ostracism, threats of violence, and actual use of force. Such acts of individual behavior can deter future defections, thus providing a public good of policing. The core of the problem, therefore, is the new dominant incentive to free-ride on the punishment and let others police defectors. Policing the police does not solve the problem for the same reason of merely shifting the issue one step further. Nevertheless, the manifest existence in the natural world of altruistic punishment—as an individual psychological response, a communal norm, and an element of institutional design—warrants examination of such behavior.

Close examination of altruistic punishment in theory and practice reveals that there are important asymmetries between classical cooperation and altruistic punishment. If individuals are willing to incur the cost of punishment, we observe the evolution of altruistic behavior in a more robust manner than in the traditional non-punishing game.[13] Hence, the first question that we have to address is whether, in fact, individuals *are* willing to punish defectors—even when such behavior is seemingly irrational in the classic economic sense.

Empirical Evidence

According to recent experimental evidence, the punishment of defectors is common in laboratory settings.[14] Up to 75 percent of subjects are willing to punish defectors even when the punishment is costly, the game is not repeated between the same players, and reputation-building is excluded, that is, other players do not observe the act of punishment. Such punishment is called altruistic since it imposes a cost on the individual while not producing any material benefit in return and while teaching the defector a lesson and making it more likely that the defector will cooperate in the future with someone else. Similar results have been achieved in political science, psychology, and economics.[15] The work by Fehr and Gachter has been especially important since the design of their experiments excluded any possibility that the punishment was actually *non*-altruistic. Recently, altruistic punishment has been also extended to explain punishment of defectors by a third party, that is, by a public good game observer and not by the [exploited] cooperator.[16]

Field studies support this experimental research. Various forms of costly self-enforcement of cooperative behavior appear to be a widespread custom in communities around the world. From fisheries to irrigation systems to grazing lands to forests and wildlife, decentralized punishment of free-riding and overexploitation is a regular institutional arrangement devised to discourage opportunistic behavior.[17]

A possible explanation of altruistic punishment is that emotions such as anger are responsible for the seemingly irrational and thus unsustainable behavior. Such an explanation, however, begs a number of questions: If we do have angry emotional responses to defection, where do they come from? If altruistic punishment decreases individual utility, how could such behavioral propensities have possibly evolved? What is so special about altruistic punishment that makes it a widespread communal custom as reflected in the field evidence, and also a strong individual propensity as reflected in the experimental research?

To answer this question we examine altruistic punishment through the three different prisms—rational choice, evolutionary adaptation, and normative evaluation. In brief, altruistic punishment is economically rational if

https://doi.org/10.1017/S1537592707071502 Published online by Cambridge University Press

pre-commitment is possible.[18] A rational individual will *have* to cooperate if the other player is pre-committed to altruistic punishment. As a result, both players are better off since no one has an incentive to free-ride and mutual cooperation is the outcome. The evolutionary basis of altruistic punishment is therefore in its self-amplifying character. Unlike cooperation, the success of altruistic punishment is *frequency dependent.* As the proportion of AP strategies increases in the population, defectors receive a progressively lower payoff. In a community where altruistic punishment becomes a wide-spread norm, defection is *least* likely. To the contrary, a community of non-punishing cooperation is *most* vulnerable to defection. Later in this article we examine economic and evolutionary properties of altruistic punishment are examined in greater detail.

Altruistic Punishment and Tit-For-Tat

Altruistic punishment may also be driven by normative considerations. Altruistic punishment as a mechanism that leads to cooperation must be normatively better than tit-for-tat (TFT), a retaliatory mechanism that entails punishment by defection. Axelrod Keohane, and other well-known political scientists describe TFT as the main reciprocity mechanism promoting cooperation.[19] This has been a standard solution to the tragedy of the commons in political science, which is surprising given the importance of normative issues for political scientists and the major deficiency of TFT as discussed below.

Although tit-for-tat can be an effective strategy leading to cooperation in theory, it is often inappropriate and ineffective in the real world.[20] In a public good game, punishing other defectors by means of one's own defection also harms individuals who cooperate. Furthermore, one's own defection in the presence of other tit-for-tat strategies leads to defection by other previously cooperating members of the community. In contrast, an altruistic punisher always cooperates for the benefit of other cooperators and punishes only defectors who deserve that. Field studies confirm that individual violations of rules lead to punishment of those who are responsible, instead of cascading defections by the rest of the group.[21] For example, overuse of a communal irrigation system leads to the punishment of the responsible individual, and not the increase of consumption by the rest of the community.[22] Similarly, underprovision of a public good—such as putting effort into a buffalo hunt—leads to the punishment of shirkers instead of shirking by the rest of the group.[23] Turnbull also shows in *The Forest People* that only the pygmies who do not contribute to a hunt are punished, often symbolically in order to be taught a lesson, while the rest of the group continue to cooperate.[24]

Theoretical treatments of altruistic punishment and closely related behaviors are numerous. In political science and economics, such examples include the theory of strong reciprocity, models of quasi-voluntary compliance, the Norms game, punishment by exit, and some of the literature on sanctions.[25] In biology, altruistic punishment is invoked by the notion of moralistic aggression, examples of punishment in animal societies, or negative reciprocity, mutual policing, repression of selfishness in the context of group selection, and so on.[26] Table 1 provides a summary of theoretical, experimental, and field research on the topic.

The table by no means represents a complete picture, but the selection should give the reader some idea about the scope of interdisciplinary interest in the question. Despite the fact that altruistic punishment in its various forms receives much attention across disciplines, scholars within each field are typically not aware about advances in other fields and disciplines. Despite addressing exactly the same problem, albeit in different domains, the gap is most striking between the studies of altruistic punishment in social and life sciences, as evident from not citing each others work (interdisciplinary works, not surprisingly, are an exception).

Classical Prisoners' Dilemma Game and Altruistic Punishment

To illustrate the logic of altruistic punishment we turn to a simple game theoretic model. Economic rationality underlying altruistic punishment can be explained through the classic Nash equilibrium solution, while the long-term properties of altruistic punishment are better examined by means of replicator dynamics and the analysis of evolutionary stability of strategies. The purpose of this model (or any model, in fact) is to simplify reality in order to sharpen our intuition and capture non-trivial aspects of the problem. How can altruistic punishment be a solution to the tragedy of the commons? The theoretical output (prediction) that we obtain is further juxtaposed with empirical reality. And, finally, normative inferences follow the discussion of empirical implications of the theoretical model.

Costly self-enforcement is not possible in the well-known Prisoners' Dilemma (PD) game, which has been a standard model predicting the tragic outcome to commons situations. In the two-player version of the game, each player (prisoner) has two available strategies: cooperate (maintain silence) and defect (confess). Confessing to the authorities is a strictly dominant strategy, making mutual defection the only equilibrium. Although the model captures the crux of the problem of collective action, it appears that certain fundamental aspects of real world dilemmas are missing—as suggested here, punishment of defectors is common when it comes to public good games. Even the prisoners in the story are likely to be aware of

https://doi.org/10.1017/S1537592707071502 Published online by Cambridge University Press

Table 1
Altruistic punishment as an integrating explanation of the puzzle of human cooperation

Theory	Experimental Research	Field Studies
Political Science - The Norms game (Axelrod 1986) - Quasi-voluntary compliance (Levi 1988) - Punishment by exit (Vanberg and Congleton 1992) **Political Economy** - Sanctions (Nossal 1989; Shavell 1987; Romer 1984) **Economics** - Repeated PD with ostracism (Hirshleifer and Rasmusen 1989) - Punishment of defectors (Binmore 1994; Sethi and Somanathan 1996) **Sociology** - Rewards and punishments (Oliver 1980) **Law** - Origin, development, and regulation of norms (McAdams 1997) - Customary International Law analogy (CIL) (Norman and Trachtman 2005) **Biology** - Reciprocal altruism and moralistic aggression (Trivers 1971) - Negative reciprocity and punishment in animal societies (Clutton-Brock and Parker 1995) - Mutual policing and repression of competition (Frank 1995; Frank 2003) **Interdisciplinary** - Evolutionary model of retribution (Boyd and Richerson 1992) - Punishment of the second-order free-riders (Henrich and Boyd 2001) - Strong reciprocity (Gintis 2000) - Computer simulation of group selection-based evolution of altruistic punishment (Boyd et al 2002) - Punishment in spatial PGG (Brandt et al. 2003) - Punishment and non-kin correlation of social traits (Gardner and West 2004) - Altruistic punishment and no play option (Fowler 2005)	**Political Science** - Enforcement of cooperation (Orbell, Van de Kragt, and Dawes 1986) - Punishment of defectors for a fee (Ostrom, Walker, and Gardner 1994) - Penalties in non-reciprocal environments (Lubell and Scholz 2001) - Egalitarian motive and altruistic punishment (Fowler et al. 2005; Dawes et al. 2007) **Experimental/Behavioral Economics** - Altruistic punishment (Fehr and Gachter 2002) - Effects of sanctions on human altruism (Fehr and Rockenbach 2003) - Third-party punishment and social norms - Punishment by rejection in the Ultimatum games (Camerer and Thaler 1995) - Reciprocity and punishment in fifteen small-scale societies (Henrich et al 2001) **Psychology/Evolutionary Psychology** - Provision of a sanctioning system as a public good (Yamagishi 1986) - Effect of the probability of punishment (Kurzban et al 2001) - Punitive sentiments (Price, Tooby, and Cosmides 2002) **Biology** - Sanctioning in plant-rhizobium mutualisms (Denison 2000; Kiers et al. 2003) - Punishment in naked model rats (Reeve 1992) and paper wasps (Reeve and Gamboa 1987)	**General studies** - Self-governance of the commons (Ostrom 1990; Ostrom, Gardner, and Walker 1994; Bromley et al 1992; Alston, Libecap, and Mueller 1999; Berkes et al 1989) - Managing small-scale fisheries (Acheson 1975; Dyer and McGoodwin 1994; Leal 1998; Berkes et al 2001; Crean and Symes 1996) - Self-governing irrigation systems (Tang 1992; Ostrom and Gardner 1993; Mabry 1996) - Grazing lands and wildlife (Netting 1981; Bromley 1992) **Specific cases** - New Zealand fisheries (Arbuckle and Metzger 2000) - Irrigation in medieval Valencia (Glick 1970) - Agroforestry by Mobisquads in Ghana (Veit et al 1991) - American Indian societies (Anderson 1995) - Order without law in cattlemen ranges, Shasta county, CA (Ellickson 1991) - California gold rush (Umbeck 1981) - Dars of Sudan (Kibreab 2002) - Spanish huertas (Maass and Anderson 1978) - Timber forest management in Nepal and Japan (Sakurai et al 2001) - Trochus management in Indonesia (Ruttan 1998)

potential retributions typical in the criminal world for not keeping silent. A simple extension of the classical PD adds another—cooperative—equilibrium to the game (see figure 1).

An altruistic punisher is a cooperator who sacrifices some of his utility in order to decrease the utility of a defection. In this game, mutual defection remains a Nash equilibrium—no player has an incentive to deviate from

https://doi.org/10.1017/S1537592707071502 Published online by Cambridge University Press

Figure 1
Prisoners' Dilemma with altruistic punishment

	Cooperator	Altruistic Punisher	Defector
Cooperator	***R, R***	***R, R***	***S, T***
Altruistic Punisher	***R, R***	***R, R***	***S-X, T-kX***
Defector	***T, S***	***T-kX, S-X***	***P, P***

Note: $T > R > P > S$, $2R > T + S$ (standard Prisoners' Dilemma assumptions), and $X > 0$, $k > 0$, where: T is the temptation payoff to free-rider, R is the reward for mutual cooperation, P is the punishment for mutual defection, S is the payoff of an exploited cooperator, X is the cost of punishment for the altruistic punisher, and kX is the cost of punishment for the defector. Typically, but not necessarily, $k > 1$. Altruistic punishment is a Nash equilibrium for $R > T - kX$.

his strategy unilaterally. However, altruistic punishment is also a Nash equilibrium if the cost of punishment for the defector is greater than the extra benefit from free-riding. The mutual defection equilibrium corresponds to the tragedy of open access resources such as the open ocean fisheries.[27] The altruistic punishment equilibrium corresponds to cooperative management of the self-governing commons (third column in table 1).

Empirical examples of successful long-term management of the commons imply that some communities manage to ensure that mutual cooperation provides greater utility than "punished free-riding"—that is to say, when following the rules of the community is more beneficial than free-riding and then being punished. Yet this may change. Field evidence also demonstrates that the cooperative equilibrium can de disrupted as a result of external factors such as natural cataclysms, infectious diseases, intergroup warfare, refugees, and state intervention. For example, droughts usually increase an individual incentive to free-ride and consume more water than others. The value of temptation is increased, and if more strict rules are not implemented, the cooperative equilibrium is in danger.[28] It should be noted, however, that the mere presence of punishment options can have a profound positive effect on the level of human cooperation. Lubell and Scholz show experimentally that in a non-reciprocal environment even small penalties for defection can lead to higher levels of cooperation.[29] This result suggests that graduated sanctions, or partial punishments, can be an effective and efficient social norm. Ellickson provides an empirical illustration of this result showing that punishment depends on both the magnitude of free-riding and the frequency of repeated offense.[30]

The cooperative equilibrium—existing due to altruistic punishment—gives both players higher utility than the mutual defection equilibrium. At the same time, altruistic punishment is weakly dominated by cooperation. In terms of individual utility, cooperation is at least as good as altruistic punishment, and sometimes (when defectors are present) even better. Ironically, cooperators in the model are, in fact, the "second-order defectors" since they do not punish defectors, that is, free-ride on the policing effort of altruistic punishers. While cooperation weakly dominates altruistic punishment, it itself is strictly dominated by defection in the absence of altruistic punishment. These dynamics raise a question whether altruistic punishment can be sustained as a social norm in the long run. To examine long-term properties of altruistic punishment we turn to evolutionary game theory.

Evolutionary Stability of Altruistic Punishment

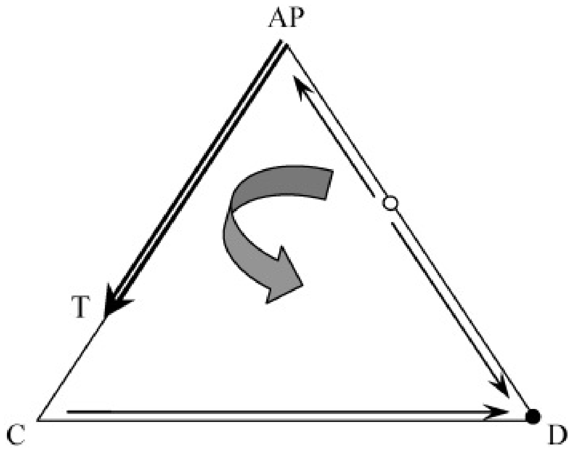
The fact that altruistic punishment is a Nash equilibrium strategy does not imply that it is also an evolutionary stable strategy (ESS). A strategy, or a type, is evolutionary stable when all members of the population adopt it and no single individual has an incentive to adopt another strategy.[31] Formally, a strategy adopted by all members of the population is called an *incumbent* strategy whereas all other strategies are called *mutant* strategies. If a mutant strategy has the same payoff as the incumbent, then the strategies are "neutrally stable."[32] The analysis of evolutionary stability allows us to examine the long-term feasibility of Nash equilibria if the game is repeated.

Defection is the only evolutionary stable strategy in the modified PD game. Nevertheless, there is still a possibility for an extended but temporary coexistence of cooperators and altruistic punishers. Intuition rightfully suggests that such a heterogeneous population cannot exist forever. In the presence of defectors, cooperation provides greater individual fitness since cooperators do not incur the cost of punishment. As a result, the proportion of altruistic punishers relative to the proportion of cooperators will be decreasing over time, even if defection is deterred. Eventually, there will be so few altruistic punishers and so many cooperators that a mutant defector will take over the population. In fact, altruistic punishment is unstable even in the absence of defectors since it is only neutrally stable against mutant cooperators. As a result the population can randomly drift away from the social norm of altruistic punishment.

The analysis of replicator dynamics confirms our intuition (see figure 2). Formally, replicator dynamics are a system of equations representing the growth of types (strategies) in the population given their relative fitness.[33]

https://doi.org/10.1017/S1537592707071502 Published online by Cambridge University Press

Figure 2
Replicator dynamics of the Prisoners' Dilemma game with altruistic punishment



Note: The corners of the simplex represent homogeneous populations of cooperators (C), defectors (D), and altruistic punishers (AP). In the AP-T region, the population is heterogeneous and consists only of cooperators and altruistic punishers. In this region, mutant defectors die out, yet cooperators have greater fitness than altruistic punishers. Below the threshold T, the proportion of cooperators becomes so large relative to the proportion of altruistic punishers that a mutant defector will have greater fitness than both cooperators and altruistic punishers and, therefore, the population will eventually converge to mutual defection (D).

In the short run, defection can be contained and the two surviving traits will be cooperation and altruistic punishment. Unfortunately for altruistic punishers and cooperators, the region AP-T is not stable since occasional defections will increase the proportion of cooperators at the expense of altruistic punishers. Eventually altruistic punishment will cease to exist as a social norm. Interestingly, the actual fitness difference between cooperators and altruistic punishers is very small. As a result the speed with which the proportion of cooperators increases relative to the proportion of altruistic punishers is also very slow (a tiny fraction of a percentage per generation). At the same time, a number of factors may easily reverse the dynamics. Examples of such factors include conformist pressure to punish defectors, multi-level selection, a "No Play" option, and a spatial structure.[34] A costless punishment solves the problem altogether by assumption.[35]

In the group selection model the key difference between cooperation and altruistic punishment becomes apparent, rendering AP as an evolutionary robust social norm. The crux of the group selection argument is that a group of cooperators will be more successful than a group of defectors in direct or indirect competition for resources. In this respect, each individual has an incentive to cooperate to make his group more successful than other groups. On the other hand, each individual also has an incentive to defect to increase his payoff within a group. It is possible to show that in the model of group selection cooperation can survive if several very strict conditions are met such as significant fitness differential between the groups, limited migration and genetic drift, and extinction and formation of new groups. The evolution of cooperation under group selection is possible but not very probable since the specific conditions above must be met simultaneously. The main difficulty is that within each group cooperation is *always* strictly dominated by defection for *all* possible parameters of the model. And this is where altruistic punishment is different: altruistic punishment is frequency dependent—the greater the proportion of punishers, the stronger is the group in the intergroup competition *and* the less beneficial is defection within a group.

Altruistic punishment retains a between-group advantage of cooperation but manages to avoid its within-group disadvantage. A homogeneous group of altruistic punishers is as successful as a homogeneous group of cooperators in the inter-group competition. At the same time, the cooperative group is much more vulnerable to within-group defection. The robustness of altruistic punishment is directly proportional to its frequency. An increase in the proportion of altruistic punishers increases the fitness of such individuals and decreases the fitness of defectors. This is different from cooperation which is frequency independent within each group where it is always dominated by defection. Thus, the groups with the higher proportions of altruistic punishers have an advantage in the between-group competition for resources and are more capable of suppressing within-group defection.

Empirical Robustness of Altruistic Punishment

The field evidence shows that altruistic punishment often underlies a solution to the problem of overexploitation of common-property resources, especially in cases when privatization and governmental control are problematic. One form of altruistic punishment is a widespread custom of "self-help", or self-enforcement of community rules by local means.[36] In this case, individuals rely on personal retaliation as a primary countermeasure against deviants. As Ellickson put it, "a measured amount of self-help—an amount that would serve to even up accounts—is the predominant and ethically preferred response."[37] Although responsibility for social control in many cases belongs to the victim, the individual has potential support from the whole group if the conflict escalates.[38] The cost of punishment generally increases with the frequency and degree of violation—producing sequentially warnings, gossip, equipment damage, threat of violence, social ostracism, and even physical harm. Such customs do not necessarily lead to anarchy and vigilante justice. Actual cases of violence

https://doi.org/10.1017/S1537592707071502 Published online by Cambridge University Press

are rare, suggesting that the threat of punishment is primarily a *deterrence* mechanism.[39] A number of empirical examples show how successful commons have established sanctions against rule violations, while the communities without established self-sanctioning mechanisms have problems in rule conformance or maintenance.[40] Interestingly, Norman and Trachtman notice that social norms along the lines of Ellickson's self-help are "roughly analogous" and "substantively similar" to the development of customary international law (CIL) in the international public setting.[41]

Four conditions appear to be necessary (but not sufficient) for altruistic punishment to be a communal norm: (1) pre-commitment to punish free-riders, (2) successful self-monitoring, (3) a common-property regime as opposed to open access, and (4) favorable legal constrains on self-enforcement.

Pre-commitment to punish defectors

Pre-commitment to punish defectors makes altruistic punishment a deterrence mechanism, not a cycle of violence. The modified Prisoners' Dilemma game that we discuss here assumes that altruistic punishment is a type along with cooperation and defection. In terms of evolutionary psychology this is a disposition triggered by appropriate environmental stimuli, notably, existence of defection. By construction, if a player's type is AP, then this player is pre-committed to punish defectors. An alternative model is a two-stage game in which players first choose whether to cooperate or defect and *then* choose whether to punish or not.[42] In this case, the game is similar to the well-known chain-store paradox, in which case pre-commitment is also important and economically rational if the game is repeated.[43] Pre-commitment is critical for the cooperative equilibrium: if players *know* that one is pre-committed to punish defection, they will not defect. This makes pre-commitment beneficial in economic terms and adaptive in evolutionary terms. What are the effective pre-commitment mechanisms?

One class of such mechanisms is psychological adaptations, such as emotions.[44] One such emotion is anger, which biologist Trivers called "moralistic aggression."[45] Although the act of punishment is irrational in an economic sense, potential defectors know that emotional response may override economic rationality. As a result, anger proves to be adaptive for *both* players: interaction with a person who is capable of anger dramatically decreases the mutual temptation to free ride. Surprisingly, another such emotion is spite, which appears to be common.[46] In public good games defectors also happen to be the highest earners in the group.[47] Therefore, they are most likely to be the targets of spite, with altruistic punishment serving merely as an instrument to reduce group inequality.

Cultural adaptations are another class of effective mechanisms allowing pre-commitment. Communities around the world have moral codes, norms, and customs that encourage not only cooperation but also punishment of unfair behavior. In fact, such practice may take an extreme form: both cooperator and defector become the object of social ostracism if the cooperator fails to punish his offender. In terms of the model, it means that cooperators who face defectors are also punished for not being altruistic punishers.[48] Pre-commitment, therefore, is a result of the communal pressure to punish defectors, which in turn is supported by the cognitive and emotional mechanisms discussed previously.

Effective self-monitoring

Effective *self-monitoring* of the commons is also necessary since undetected cheating makes altruistic punishment irrelevant. In terms of the model, the probability of getting away with defection is functionally the same as adding a weight—less than one—to the cost of punishment. In the environment where defection is completely unobservable, altruistic punishment becomes identical to cooperation, and we return to the classical version of the Prisoners' Dilemma game. Trivers, for example, argued that defectors could benefit from "subtle cheating" helping them avoid repercussions for their behavior; in response, however, humans have evolved cheater-detection cognitive adaptations.[49]

This argument is supported by research in evolutionary psychology.[50] Cheater-detection cognitive adaptations indicate that humans are good at spotting free-riding. In particular, experiments on Wason selection task demonstrate that humans are capable of making complex logical inferences in a social environment where cheating is possible. Such cognitive adaptations, however, are only effective for small groups of individuals (less than 100) and in the absence of any sophisticated tools that can be used for cheating. These psychological mechanisms are believed to have evolved during the environment of evolutionary adaptedness (Pleistocene), or 99.9% of the human evolutionary development. In the contemporary environment, different from the Pleistocene and often characterized by big group sizes and new technologies, cheater-detection adaptations may be ineffective. Examination of the collective action problem through the prism of group size is similar to Mancur Olson's classic analysis but the underlying economic and evolutionary arguments are different.[51] According to the standard economic view, individuals in large groups are least likely to contribute because the benefit is diluted among members of the group. In the evolutionary framework, individuals are less likely to contribute because it is easier to get away with free-riding in a large group—a consequence of the nature of the environment in which human cognitive apparatus had evolved.

In response to the natural limits of cognitive cheater-detection, communities around the world have evolved sophisticated monitoring mechanisms designed to prevent cheating.[52] Although new technologies make it easier to cheat, they also make it easier to monitor and detect the violation of rules.

Common-property regime vs. open access

Existence of such communal rules is possible and appropriate for common-property regimes as opposed to open access.[53] Common-property regimes are fundamentally different from open access insofar as the former have clearly defined membership status, community borders, established norms which are common knowledge, entry barriers for outsiders, and explicit or implicit community rules.[54] Hardin's tragedy of the commons describes the tragedy of open access resources—not the commons—when there are no established norms and rules, when monitoring and sanctioning of others is difficult.[55] In terms of the theoretical model, it means that the punishment of defectors may be impossible or even inappropriate; therefore, management of open access resources is more challenging. As a consequence, one of the policy implications is that altruistic punishment can be a solution of the tragedy of open access resources only if the latter become common property.

Legal constraints

Legal constraints typically promote cooperation through the state-based law enforcement. Under many circumstances state intervention proves to be necessary and beneficial. Interestingly, it may also have a surprisingly negative effect on commons management especially in less-developed countries.[56] Max Weber argued that the modern state seeks "to monopolize the legitimate use of force."[57] Monopolization of force, however, undermines decentralized punishment by making it illegal. Frohlich and Oppenheimer also showed that existence of various institutional constraints and incentive compatible devices in public good games can actually decrease voluntary giving by eliminating ethical concerns.[58]

If the state alternatives to self-enforcement are not effective, the cooperative equilibrium is in danger. Field studies document how centralized intervention can fail to preserve local customs and lead to the tragedy of the commons.[59] In the Sudan as well as some other developing countries, the state rather than communal ownership is seen as the "major cause of inappropriate land use practices and consequently depletion of [common-pool resources]."[60] If altruistic punishment plays an important role in successful management of the commons, state intervention may create more problems than benefits. As Vernon Smith put it in his Noble Prize address,

> voluntary private associations for sharing the cost of a common good—policing—were subsequently undermined by statehood, and the publicly financed local sheriff as the recognized monopoly law enforcement officer. This observation contradicts the myth that a central function of government is to "solve" the free-rider problem in the private provision of public goods.[61]

Breaking the communal self-enforcement rules and norms could entail risk whereas re-creation of self-governing commons is a difficult task.[62]

In short, available empirical evidence suggests that precommitment, successful monitoring, common-property regime, and a certain degree of independence are necessary attributes of the successful self-enforcement of cooperation in the commons. An interesting link between the two factors was suggested by Elinor Ostrom, who reports a large number of empirical cases showing that "overexploitation of common-pool resources occurred when open access prevailed either because no set of individuals had property rights or because state property was treated as open-access property."[63]

Conclusion

It has been a folk theorem that punishment can sustain cooperation. In the traditional rational choice models of iterated play, punishment means retaliatory defection. A possibility of continuing cooperation and punishment by other means has been largely ignored by the general political science audience, and by policy makers, because costly punishment can be seen as a public good itself, thus, subject to the same problem of individual free-riding. Nevertheless, experimental evidence and field studies suggest that altruistic punishment is common. Individuals are willing to incur a cost in order to punish defectors. Theoretical examinations of the phenomenon can be found in various disciplines: political science, economics, biology, computer science, psychology, and others. The overlap between the study of public goods in the social sciences and altruistic behavior in the life sciences suggests altruistic punishment as a possible solution to the tragedy of the commons. Costly self-enforcement of cooperation appears to be a part of human psychological apparatus as well as communal rules and norms. AP is rational in economic terms if precommitment is possible. It is also robust in evolutionary terms, especially in comparison with non-punishing cooperation. Unlike cooperation, altruistic punishment is frequency dependent: the higher the proportion of altruistic punishers in the population the greater is their fitness. Although altruistic punishment is not asymptotically stable within a single group, the problem can be offset by a within-group conformist pressure or between-group competition. In the context of group selection, altruistic punishment retains the strength of cooperation in between-group competition and, at the same time, prevents defection from taking over within the group. In normative terms, altruistic punishment is a more appealing norm than punishment by *defection.* In the commons, Tit-for-Tat and other trigger strategies not only punish defectors but also harm individuals who cooperate, which

https://doi.org/10.1017/S1537592707071502 Published online by Cambridge University Press

may lead to a new wave of defections. Altruistic punishers always cooperate and punish only those who deserve it. Interestingly, a by-product of altruistic punishment is improved group equality since defectors also happen to be the highest earners in public good games.

A possibility of altruistic punishment as a solution to the tragedy of the commons has immediate policy implications. If cheating can be prevented, if there is no emergency, and if the commons are characterized by the common-property regime, the community may be successful without external assistance such as state intervention. On the other hand, if monitoring is costly or impossible, if the commons are characterized by the open access regime, and if external factors such natural cataclysms are present, external assistance may be necessary. Failure to differentiate between the two cases as well as failure to recognize the importance of self-enforcement of cooperation as a community norm will only accelerate the opportunistic behavior and overexploitation of available resources. The tragedy of the commons often happens when individuals start treating their common-property resources as open access property. Recognizing altruistic punishment as a vital attribute of the common-property management may explain some of the counter-productive policies. When the state monopolizes the use of force it automatically undermines the mechanisms of internal enforcement of cooperation. External enforcement may or may not solve the free-rider problem, but it is certain to bring the commons one step closer to being an open access regime by taking away the social norm of altruistic punishment.

This argument implies that to solve the tragedy of the commons, policy-makers should not necessarily get rid of the *commons* by means of privatization or centralized coercion. To the contrary, to prevent the tragedy, policy-makers may want to re-create and reinforce the commons as a common-property regime with a certain degree of sovereignty, characterized by its customs and norms. This conclusion may be potentially relevant not only to the governance of common-property regimes, but more broadly to the general issue of evolved local institutions and government control.

Notes

1 Hamilton 1964; Trivers 1971.
2 Flood 1952; Olson 1965.
3 Hobbes 1947.
4 Hardin 1968.
5 Plott 1983; Isaac, Walker, and Thomas 1984; Orbell, Van de Kragt, and Dawes 1988; Ledyard 1995; Lubell and Scholz 2001.
6 Ostrom 1990; Ostrom, Gardner, and Walker 1994.
7 Kreps et al. 1982; Fudenberg and Maskin 1986; Binmore 1992.
8 Hamilton 1964, 1975; Trivers 1971; Alexander 1987; Sober and Wilson 1998.
9 Maynard Smith 1982; Fowler 2005; Boyd and Richerson 1992; cf. Axelrod 1997; Nowak and Sigmund 1998; Orbell et al. 2004.
10 Nowak 2006.
11 Smith 1981; Anderson and McChesney 2003; Ophuls 1973.
12 Elster 1989.
13 Cf. Boyd et al. 2003; Boyd and Richerson 1992; Henrich and Boyd 2001.
14 Fehr and Gachter 2002.
15 Orbell, Van de Kragt, and Dawes 1988; Yamagishi 1986; Ostrom, Gardner, and Walker 1994; Price, Cosmides, and Tooby 2002; Fehr and Gachter 2002.
16 Fehr and Fischbacher 2004.
17 On fisheries see Acheson 1975; Dyer and McGoodwin 1994; Crean and Symes 1996; Leal 1996; Berkes et al. 2001. On irrigation systems see Tang 1992; Ostrom and Gardner 1993; Mabry 1996. On grazing lands see Netting 1981; Ellickson 1991; Anderson 1995. On forests and wildlife see Bromley 1992; Kibreab 2002. See also Ostrom 1990 and Ostrom, Gardner, and Walker 1994 for a comprehensive overview of self-governing commons.
18 Cf. chain store paradox as in Kreps and Wilson 1982; according to the chain store paradox a monopoly may commit itself to punishment of all entrants into the market in order to build reputation and deter future invasions.
19 Axelrod 1984; Keohane 1986.
20 Axelrod and Hamilton 1981.
21 Ostrom 1990; Ellickson 1991; Anderson 1995.
22 Tang 1992; Mabry 1996.
23 Terry L. Anderson, personal communication 2003.
24 Turnbull 1961.
25 On the theory of strong reciprocity see Sethi and Somanathan 1996; Gintis 2000. On models of quasi-voluntary compliance see Levi 1988. On the Norms game see Axelrod 1986. On punishment by exit see Vanberg and Congleton 1992. For the literature on sanctions see Romer 1984, Shavell 1987, and Nossal 1989.
26 On the notion of moralistic aggression see Trivers 1971. On punishment in animal societies or negative reciprocity see Clutton-Brock and Parker 1995. On mutual policing see Frank 1995. On repression of selfishness in the context of group selection see Sober and Wilson 1998.
27 Meltzer 1994.
28 Ostrom 1990; Tang 1992; Mabry 1996.
29 Lubell and Scholz 2001.
30 Ellickson 1991.
31 Maynard Smith 1982.
32 Samuelson 1997.

33 Maynard Smith 1982, Samuelson 1997.
34 On conformist pressure to punish defectors see Henrich and Boyd 2001 and Axelrod 1986. On multi-level selection see Sober and Wilson 1998. On "No Play" option see Fowler 2005. On spatial structure see Brandt et al. 2003.
35 Hirshleifer and Rasmusen 1989; McAdams 1997.
36 Ellickson 1991.
37 Ellickson 1991, 57.
38 Anderson 1995.
39 Ellickson 1991; Berkes et al. 2001.
40 Tang 1992.
41 Norman and Trachtman 2005, 545.
42 Cf. Axelrod 1986, Sethi and Somanathan 1996.
43 Kreps and Wilson 1982.
44 Tooby and Cosmides 1992.
45 Fehr and Gachter 2002, Trivers 1971.
46 Dawes et al. 2007.
47 Fowler, Johnson, and Smirnov 2005.
48 Cf. Boyd and Richerson 1992.
49 Trivers 1971.
50 Tooby and Cosmides 1992, Gigerenzer and Hug 1992, Orbell et al. 2004.
51 Olson 1965.
52 Ostrom 1990, Bromley 1992, Tang 1992, Crean and Symes 1996, Alston, Libecap, and Mueller 1999.
53 Ciriacy-Wantrup and Bishop 1975.
54 Stevenson 1991.
55 Hardin 1968. Meltzer 1994, Berkes et al. 1989, Kibreab 2002.
56 Kibreab 2002, Smith 2003.
57 Weber 1958.
58 Frohlich and Oppenheimer 1995. I am grateful to John Orbell for bringing this to my attention.
59 Berkes et al. 2001, Tang 1992.
60 Kibreab 2002, 403.
61 Smith 2003, n. 52.
62 Bromley 1992, Mabry 1996, Berkes et al. 2001, Kibreab 2002.
63 Ostrom 1992, 312.

References

Acheson, J. M. 1975. The lobster fiefs: Economic and ecological effects of territoriality in the Maine lobster industry. *Human Ecology* 3 (3): 183–207.

Alexander, R. D. 1987. *The Biology of Moral Systems.* Aldine de Gruyter: New York.

Alston, Lee J., Gary D. Libecap, and Bernardo Mueller. 1999. *Titles, Conflict, and Land Use.* Ann Arbor: University of Michigan Press.

Anderson, Terry. L. 1995. *Sovereign Nations or Reservations?* San Francisco: Pacific Research Institute for Public Policy.

Anderson, Terry L., and Fred S. McChesney. 2003. *Property Rights: Cooperation, Conflict, and Law.* Princeton: Princeton University Press.

Arbuckle, M., and M. Metzger. 2000. *Food for Thought: A Brief History of the Future of Fisheries Management.* Nelson, New Zealand: Challenger Scallop Enhancement Company Limited.

Axelrod, Robert. 1984. *The Evolution of Cooperation.* Basic Books.

———. 1986. An evolutionary approach to norms. *American Political Science Review* 80 (4): 1095–111.

———. 1997. *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration.* Princeton, NJ: Princeton University Press, 1997.

Axelrod, R., and W.D. Hamilton. 1981. The evolution of cooperation. *Science* 211: 1390–96.

Berkes, F., D. Feeny, B.J. McCay, and J.M. Acheson. 1989. The benefits of the commons. *Nature* 340: 91–93.

Berkes, F., R. Mahon, P. McConney, R. Pollnac, and R. Pomeroy. 2001. *Managing Small-scale Fisheries.* Ottawa: International Development research Centre.

Binmore, Ken. 1992. *Game Theory and Social Contract. Volume 2: Just Playing.* Lexington: D.C. Heath.

Boyd, Robert, Herbert Gintis, Samuel Bowles, and Peter J. Richerson. 2003. The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences USA,* 100 (6): 3531–3535.

Boyd, R., and P.J. Richerson. 1992. Punishment allows the evolution of cooperation (or anything else) in sizeable groups. *Ethoogy and Socibiology* 13: 171–95.

Brandt, Hannelore, Christoph Hauert, and Karl Sigmund. 2003. Punishment and reputation in spatial public goods games. *Proceedings: Biological Sciences* 270 (1519): 1099–104.

Bromley, D. W., ed. 1992. *Making the Commons Work.* San Francisco: Institute for Contemporary Studies.

Ciriacy-Wantrup, Siegfried V., and Richard C. Bishop. 1975. Common property as a concept in natural resource policy. *Natural Resources Journal* 15: 713–27.

Clutton-Brock, T. H., and G. A. Parker. 1995. Punishment in animal societies. *Nature* 373: 209–16.

Crean, Kevin, and David Symes, eds. 1996. *Fisheries Management in Crisis.* Oxford: Fishing News Books.

Dawes, Christopher, Tim Johnson, James Fowler, Richard McElreath, and Oleg Smirnov. 2007. *Nature* 446: 794–96.

Dyer, Christopher L., and James R. McGoodwin, eds. 1994. *Folk Management in the World's Fisheries: Lessons for Modern Fisheries Management.* Niwot: University Press of Colorado.

Ellickson, Robert C. 1991. *Order without Law: How Neighbors Settle Disputes.* Cambridge: Harvard University Press.

https://doi.org/10.1017/S1537592707071502 Published online by Cambridge University Press

Elster, Jon. 1989. *The Cement of Society: A Study of Social Order.* Cambridge: Cambridge University Press.

Fehr, Ernst, and Urs Fischbacher. 2004. Third-party punishment and social norms. *Evolution and Human Behavior* 25: 63–87.

Fehr, Ernst, and Simon Gachter. 2002. Altruistic punishment in humans. *Nature* 415: 137–40.

Fehr, Ernst, and B. Rockenbach. 2003. Detrimental effects of sanctions on human altruism. *Nature* 422: 137–40.

Field, Alexander J. 2001. *Altruistically Inclined? The Behavioral Sciences, Evolutionary Theory, and the Origins of Reciprocity.* Ann Arbor: University of Michigan Press.

Flood, M.M. 1952. Some experimental games. Research memorandum RM-789. Santa Monica, CA: RAND Corporation.

Fowler, James. 2005. Altruistic Punishment and the Origin of Cooperation. *Proceedings of the National Academy of Sciences USA* 109 (19): 7047–7049.

Fowler, James, Tim Johnson, and Oleg Smirnov. 2005. Egalitarian motive and altruistic punishment. *Nature* 433: doi:10.1038/nature03256 (06 January).

Frank, Steven A. 1995. Mutual policing and repression of competition in the evolution of cooperative groups. *Nature* 377: 520–522.

———. 2003. Perspective: Repression of competition and the evolution of cooperation. *Evolution* 57 (4): 693–705.

Frohlich, Norman, and Joe A. Oppenheimer. 1995. The incompatibility of incentive compatible devices and ethical behavior: Some experimental results and insights. *Public Choice Studies* 25: 24–51.

Fudenberg, Drew, and Eric Maskin. 1986. The Folk Theorem in repeated games with discounting and incomplete information. *Econometrica* 54: 533–54.

Gardner, Andy, and Stuart A. West. 2004. Cooperation and punishment, especially in humans. *American Naturalist* 164 (6): 753–64.

Gigerenzer, Gerd, and K. Hug. 1992. Domain specific reasoning: Social contracts, cheating, and perspective change. *Cognition* 43: 127–71.

Gintis, H. 2000. Strong reciprocity and human sociality. *Journal of Theoretical Biology* 206: 169–79

Glick, T. F. 1970. *Irrigation and Society in Medieval Valencia.* Cambridge: Harvard University Press.

Hamilton, W. D. 1964. Genetical evolution of social behavior I and II. *Journal of Theoretical Biology* 7: 1–52.

———. 1975. Innate social aptitudes of man: An approach from evolutionary genetics. In *Biosocial Anthropology,* ed. R. Fox. New York: Wiley.

Hardin, G. 1968. The tragedy of the commons. *Science* 162: 1243–48.

Henrich, Joseph, and Robert Boyd. 2001. Why people punish defectors. *Journal of Theoretical Biology* 208: 79–89.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath. 2001. Cooperation, reciprocity and punishment in fifteen small-scale societies. *American Economic Review* 91: 73–8.

Hirshleifer, D., and E. Rasmusen. 1989. Cooperation in a repeated prisoner's dilemma with ostracism. *Journal of Economic Behavior and Organization* 12: 87–106.

Hobbes, Thomas. 1947 [1651]. *Leviathan.* New York: E. P. Dutton and Co.

Isaac, R. M., J. Walker, and S. Thomas. 1984. Divergent evidence on free riding: An experimental examination of possible explanations. *Public Choice* 43: 113–49.

Keohane, Robert O. 1986. Reciprocity in international relations. *International Organization* 40 (1): 1–27.

Kibreab, G. 2002. *State Intervention and the Environment in Sudan, 1889–1989.* Lewiston, NY: Edwin Mellen Press.

Kreps, David, and Robert Wilson. 1982. Reputation and imperfect information. *Journal of Economic Theory* 27: 253–79.

Kreps, David, Paul Milgrom, John Roberts, and Robert Wilson. 1982. Rational cooperation in the finitely repeated Prisoners' Dilemma. *Journal of Economic Theory* 27: 245–52.

Kurzban, R., K. McCabe, K. Smith, and B. Wilson. 2001. Incremental Commitment in a Real-time Public Goods Game. *Personality and Social Psychology Bulletin* 27: 1662–73.

Leal, Donald R. 1996. Community-Run Fisheries: Avoiding the Tragedy of the Commons. *PERC Policy Series* 7. Bozeman, MT: PERC.

Ledyard, J. 1995. Public goods: A survey of experimental research. In *The Handbook of Experimental Economics,* ed. J.H. Kagel and A.E. Roth. Princeton: Princeton University Press.

Levi, Margaret. 1988. *Of Rule and Revenue.* Berkeley: University of California Press.

Lubell, Mark, and John T. Scholz. 2001. Cooperation, reciprocity, and the collective-action heuristic. *American Journal of Political Science* 45: 160–78.

Maas, Arthur, and Raymond L. Anderson. 1978. *And the Desert Shall Rejoice: Conflict, Growth, and Justice in Arid Environment.* Cambridge: MIT Press.

Mabry, J. B., ed. 1996. *Canals and Communities.* Tucson: University of Arizona Press.

Maynard Smith, John. 1982. *Evolution and the Theory of Games.* Cambridge: Cambridge University Press.

McAdams, R.H. 1997. The origin, development, and regulation of norms. *Michigan Law Review* 96 (2): 338–433.

Meltzer, E. 1994. Global overview of straddling and highly migratory fish stocks: The nonsustainable nature of high seas fisheries. *Ocean Development and International Law* 25: 255–344.

Netting, R. McC. 1981. *Balancing on an Alp: Ecological Change and Continuity in a Swiss Mountain Community.* Cambridge: Cambridge University Press.

Norman, George, and Joel P. Trachtman. 2005. The Customary International Law Game. *American Journal of International Law* 99 (3): 541–80.

Nossal, Kim R. 1989. International sanctions as international punishment. *International Organization* 43: 301–32.

Nowak, M.A. 2006. *Evolutionary Dynamics.* Cambridge, MA: Belknap Press.

Nowak, M.A., and K. Sigmund. 1998. Evolution of indirect reciprocity by image scoring. *Nature* 393: 573–77.

Oliver, P. 1980. Rewards and punishments as selective incentives for collective action: theoretical investigations. *American Journal of Sociology* 85: 1365–75.

Olson, Mancur. 1965. *The Logic of Collective Action.* Cambridge: Harvard University Press.

Ophuls, W. 1973. Leviathan or oblivion. In *Toward a Steady State Economy,* ed. H. E. Daly. San Francisco: Freeman.

Orbell, John, Tomonori Morikawa, Jason Hartwig, James Hanley, and Nicholas Allen. 2004. Machiavellian intelligence, rationality, and the evolution of cooperative dispositions. *American Political Science Review* 98 (1): 1–15.

Orbell, J. M., A. van de Kragt, and R. Dawes. 1988. Explaining discussion-induced cooperation. *Journal of Personality and Social Psychology* 54: 811–19.

Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action.* Cambridge: Cambridge University Press.

———. 1992. The rudiments of a theory of the origins, survival, and performance of common-property institutions. In *Making the Commons Work,* ed. D. W. Bromley. San Francisco: ICS Press.

Ostrom, Elinor, and Roy Gardner. 1993. Coping with asymmetries in the commons: Self-governing irrigation systems can work. *Journal of Economic Perspectives* 7: 93–112.

Ostrom, Elinor, Roy Gardner, and James Walker. 1994. *Rules, Games, and Common-Pool Resources.* Ann Arbor: University of Michigan Press.

Plott, Charles R. 1983. Externalities and corrective policies in experimental markets. *Economic Journal* 93: 106–27.

Price, M.E., L. Cosmides, and J. Tooby. 2002. Punitive sentiment as an anti-free rider psychological device. *Evolution and Human Behavior* 23: 203–31.

Reeve, H. K. 1992. Queen activation of lazy workers in colonies of the eusocial naked mole-rat. *Nature* 358: 147–49.

Reeve, H.K., and J. Gamboa. 1987. Queen regulation of worker foraging in paper wasps: a social feedback-control system (*Polistes fuscatus, Hymenoptera, Vespidae*). *Behavior* 102: 147–67.

Romer, David. 1984. The theory of social custom: A modification and some extensions. *Quarterly Journal of Economics* 99: 717–27.

Ruttan, L. 1998. Closing the commons: Cooperation for gain or restraint? *Human Ecology* 26: 43–66.

Sakurai, T., Y. Kijima, R. Pokharel, S. Rayamajhi, and K. Otsuka. 2001. Timber forest management in Nepal and Japan. In *Land Tenure and Natural Resource Management: A Comparative Study of Agrarian Communities in Asia and Africa,* ed. K. Otsuka and F. Place. Baltimore: Johns Hopkins University Press.

Samuelson, Larry. 1997. *Evolutionary Games and Equilibrium Selection.* Cambridge: MIT Press.

Sethi, R., and E. Somanathan. 1996. The evolution of social norms in common property resource use. *American Economic Review* 86: 766–88.

Shavell, Steven. 1987. The optimal use of nonmonetary sanctions as a deterrent. *American Economic Review* 77: 584–92.

Smith, R. J. 1981. Resolving the tragedy of the commons by creating private property rights in wildlife. *CATO Journal* 1: 439–68.

Smith, Vernon. 2003. Constructivist and ecological rationality in economics. *American Economic Review* 93: 465–508.

Sober, E., and D.S. Wilson. 1998. *Unto Others: The Evolution and Psychology of Unselfish Behaviour*. Cambridge, MA: Harvard University Press.

Stevenson, Glenn G. 1991. *Common Property Economics: A General Theory and Land Use Applications.* Cambridge: Cambridge University Press.

Tang, S. Y. 1992. *Institutions and Collective Action.* San Francisco: Institute for Contemporary Studies.

Tooby, John, and Leda Cosmides. 1992. The psychological foundations of culture. In *The Adapted Mind: Evolutionary Psychology and the Generation of Culture,* ed. Jerome H. Barkow, Leda Cosmides, and John Tooby. New York: Oxford University Press.

Trivers, Robert. 1971. The evolution of reciprocal altruism. *Quarterly Review of Biology* 46: 35–57.

Turnbull, Colin Macmillan. 1961. *The Forest People.* New York: Simon and Schuster.

Umbeck, John R. 1981. *A Theory of Property Rights: With Application to the California Gold Rush.* Ames: Iowa State University Press.

Vanberg, V.J. and R.D. Congleton. 1992. Rationality, morality and exit. *American Political Science Review* 86 (2): 418–31.
Veit, Peter G., Clement Dorm-Adzobu, and Okyeame Ampadu-Agyei. 1991. *Community Institutions in Resource Management: Agroforestry by Mobisquads in Ghana.* World Resources Institute.
Weber, M. 1958. *From Max Weber: Essays in Sociology.* Oxford: Oxford University Press.
Yamagishi, T. 1986. The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology* 51: 110–16.