Signals, Symbols, and Human Cooperation

T. K. Ahn
Department of Political Science
531a Bellamy Building
Florida State University
Tallahasse, FL 32306
(850) 644-4540

Marco A. Janssen
Center for the Study of Institutions, Population, and Environmental Change
Indiana University
408 North Indiana Avenue
Bloomington, IN 47408-3799
(812) 855-5178
Fax (812) 855-2634
maajanss@indiana.edu

Elinor Ostrom
Workshop in Political Theory and Policy Analysis
Center for the Study of Institutions, Population, and Environmental Change
Department of Political Science
Indiana University
513 North Park Street
Bloomington, IN 47408
(812) 855-0441
Fax (812) 855-3150
ostrom@indiana.edu

Signals, Symbols, and Human Cooperation

T.K. Ahn, Marco A. Janssen, and Elinor Ostrom

Human sociality differs from that of other mammals in that only humans have generated societies whose complexity approaches and eventually surpasses that of social insects and colonial invertebrates (Wilson 2000[1975]). Within complex human societies individuals engage in a wide diversity of cooperative actions leading to joint outcomes. Many have studied how this level of cooperation has emerged in an evolutionary process based on competition.

Various contributions in this book argue that the direct benefits of cooperation may be sufficient to maintain cooperative relationships (see also Clutton-Brock 2002). Direct benefits of cooperation, however, while relevant in explaining cooperative hunting or cooperative breeding, are less powerful in explaining cooperation in complex societies that have evolved during the last several millennia. In many instances, cooperation produces indirect benefits over time rather than immediate returns essential for physical survival.

In fact, we can distinguish two classes of cooperation: (1) where a temptation to defect exists because the individual contributes more than it gains from its own contribution, and (2) where no temptation to defect exists. The second class of cooperation exists in processes where the sheer number of organisms acting together generates fitness advantages. In the terminology of modern economics, the second type of cooperation problem that individual organisms face is called a *coordination* problem. The

need for coordination to take advantage of group size is an important source for the evolution of sociality among humans and animals. In a coordination process everyone benefits more from cooperating with others than they contribute.

While the second type of cooperation is largely explained by the principle of evolutionary adaptation between individuals and their evolutionary environment, the evolution of the first kind of cooperation is harder to explain. The Darwinian evolutionary principle assumes that organisms that do not maximize fitness will be weeded out by the force of evolution. At least on the surface, however, the kind of behavior that does not seem to maximize individual fitness is exactly the behavior that is needed for cooperation of the first kind to exist. The core question is why would an individual who contributes more to others than it receives survive in a competitive process. Even though the individual is better off when in a group of cooperating individuals than when in a group of non-cooperating individuals, the individual maximizes short-term returns when others cooperate and the individual defects.

This chapter concerns the first type of cooperation among humans in situations in which the temptation to defect exists. In particular, we address how effective signals and symbols evolve to facilitate cooperation. One of the main puzzles in human societies is why costly cooperation is frequent among genetically unrelated people, in non-repeated interactions, and in the contexts in which gains from reputation are small or absent (Fehr and Gächter 2002). We argue that the ability of humans to use signals and craft symbolic systems facilitates cooperation in non-repeated interactions and stimulates the development of complex social organizations. This symbolic capability of humans is the key that differentiates them from nonhuman animals. Over time, the use of artificial

symbols to establish, to convey, and to detect reputation, has brought forth the possibility of human cooperation on unprecedented scales.

Throughout the rest of this chapter, we focus on cooperation of the first type -- on behavior that, at least in the short term, does not appear to be fitness maximizing or incentive-compatible. Cooperation can be viewed as a subcategory of altruism (as it is typically understood among evolutionary biologists as a kind of behavior). We use cooperation in a multilateral context. That is, in a two-organism interaction, for example, both participants should have the opportunity to confer payoff (fitness or welfare) to each other. The Prisoner's Dilemma is the most famous example of the situation in which cooperation of this type may or may not exist. In a Prisoner's Dilemma, two players can cooperate or defect. The payoff matrix of their choices provides the individual highest payoff when a player defects while the opponent cooperates. Selfish rational players will therefore defect, while both will be better off when they both cooperate.

Among animals, cooperation among non-kin has been mainly understood in terms of Trivers' (1971) theory of reciprocal altruism. Reciprocal altruism requires repeated interactions among individuals with the capability of recognizing each other's genetic programming or past behavior. Examples of cooperation based on type-detection and memory of past behavior are observed among fish, vampire bats, and chimpanzees (de Waal 1989, 1997; Kurzban 2003). Evolutionary game models of repeated Prisoner's Dilemma (Axelrod 1981; Axelrod and Hamilton 1981) provide formal theories that support the evolution of cooperation generated by organisms using Tit-for-Tat type strategies. Later studies by Boyd and Lorberbaum (1987), Lorberbaum (1994), and Bendor and Swistak (1997) show that Tit-for-Tat, or any other pure or mixed strategy, is

not by itself ultimately stable. That is, there are always possible combinations of strategies that can invade a population composed of a single strategy. However, the researchers also note that, in terms of relative stability, Tit-for-Tat style strategies have the best chance of evolving.

Among animals, it should be noted, the evidence of the supposed reciprocal cooperation is not as strong as the theories suggest (Stephens et al. 2002). The main reason seems to be the high discount rates of animals for whom surviving today is frequently what counts the most. In an ingenious repeated prisoner's dilemma experiment using birds as experimental subjects, Stephens et al. (2002) find that only when a low discount rate is artificially induced do Blue Jays respond in a manner consistent with the Tit-for-Tat strategy. In sum, cooperation among unrelated animals is rare even in repeated situations when substantial fitness benefits from defection exist.

Humans, on the other hand, show remarkably different cooperative patterns. Cooperation based on the reciprocity principle in repeated situations is ubiquitous, although not universal. Even when a prisoner's dilemma is repeated in an experimental laboratory for a clearly pre-announced number of rounds, human subjects sustain cooperation until near the end of repeated game (for example, Selten and Stoeker 1986; Isaac and Walker 1988; Andreoni and Miller 1993; Schmidt et al. 2001). While these experiments show more cooperation than theories based on rational and egoistic individuals would predict, there is also evidence that infinite repetition does not necessarily guarantee universal cooperation.

One possible explanation for more than predicted levels of cooperation in finitely repeated Prisoner's Dilemmas but less than full, though still substantial, cooperation in

infinitely repeated Prisoner's Dilemma is Frank's (1988) account of the role of emotions. Frank argues that a substantial proportion of humans are emotionally committed to reciprocal cooperation. That is, they feel good when mutual cooperation is achieved and feel bad when defecting on cooperative partners or when others defect on them. This hypothesis is recently supported by neuroscientific research. Rilling et al. (2002) performed iterated Prisoner's Dilemma Games, while one of the subjects was connected to a MRI machine. They found that mutual cooperation was associated with consistent activation in brain areas that have been linked with reward processes. On the other hand, humans, similar to some extent to the Blue Jays in Stephen et al.'s experiments, do not necessarily have the level of prudence required to resist temptation to defect even when defection is not a rational payoff-maximizing behavior. Therefore, the distinction between finite repetition and infinite repetition may be less useful in practice than it is viewed in economic theories. The preference for fair outcomes supported by emotions provides a consistent explanation for cooperation in single-shot or finitely repeated Prisoner's dilemma as well as the less-than-full cooperation in potentially long-term relationships.

Another characteristic of human cooperation that differs from animal cooperation is its scope, complexity, and flexibility. While eusocial insects also show fairly complicated social structures, their cooperation is mainly limited to genetically related individuals and to predictable patterns. Kinship still defines an important pattern of cooperation in humans. Human cooperation in modern times is, however, in spite of apparent similarities, qualitatively different from cooperation based on kinship.

First of all, at the individual level, the *scope* of partners with whom an individual cooperates expands far beyond kin or any genetically defined boundaries. Humans cooperate with strangers and build long-term relationships from scratch with non-kin. Even more distinctive, many humans cooperate with strangers with whom there is not much prospect for building lasting relationships. For example, during holidays or business trips, people give the waitress a tip in a restaurant to which they will never return. There is evidence that group-level differences in economic organization and the structure of social interactions explain a substantial portion of the behavioral variation across societies. In a study of 15 small-scale societies, economists and anthropologists found that the higher the degree of market integration and the higher the payoffs to cooperation in everyday life, the greater the level of prosociality expressed in experimental games (Henrich et al. 2001).

Large *scale* is another uniquely human characteristic of cooperation, except for in kin-based insect colonies. While among nonhuman primates, cooperation is limited to dyadic relationships or small groups, humans, via complex social organizations, have achieved scales of cooperation reaching thousands and even millions of individuals. How do humans achieve cooperation of such scope and scale? That is the central question of this chapter. We answer the question by examining the roles of signals and symbols in human cooperation. The scope and scale of human cooperation has been made possible by the fact that a large proportion of, though not all, humans do not maximize fitness (biology) or material wealth (economics). In the presence of a substantial number of other humans, who are fitness (wealth) maximizers, how these non-maximizers survive the evolutionary processes is a major puzzle.

Our answer to this puzzle derives from two sources: signals and symbols. For humans, the evolved biological signals are grounded on neurological processes and, thus, not easy to fake. Further the ability of humans to detect those signals and to behave contingently based on the detected signals is the fundamental biological mechanism that supports human cooperation at a remarkable scope and scale. Further, cultural and social development creates secondary signals, some of which are properly called symbols.

For those sharing a culture, cultural symbols frequently define what constitutes cooperation when there are uncertainties regarding the meaning of actions and inform people how to coordinate when there are many different ways of cooperating. Reputational symbols help individuals to detect one another and, thus, potential behavior during non-frequent encounters in situations where biological signals are not available or are unreliable.

SIGNALS

Signals are a means of communication. Wilson (2000[1975]: 176) defines biological communication as "the action on the part of one organism (or cell) that alters the probability pattern of behavior in another organism." In this definition of communication, signals are the bits of information that emanate, voluntarily or involuntarily, from the sender and reach the receiver in ways that affect the receiver's behavior. The role of biological signals in the evolution of human cooperation is discussed intensively by Frank (1988) who draws on Darwin (1873[1872]) and modern research on facial expression.

Imagine an individual in a one-shot *sequential* Prisoner's Dilemma situation. Game-theoretic analysis of the situation based on an assumption that individuals are all selfish maximizers predicts that any individual will defect. Regardless of whether one is a first or second mover, the second obtains a larger payoff by defecting. The individual, knowing this, is predicted to defect.

The available evidence does not strongly support this prediction. For example, in a set of three single-shot experiments using the same experimental protocol in three countries, Japan, Korea, and the United States (see Ahn et al., in press), between one half and two thirds of second movers cooperate in single-shot sequential Prisoner's Dilemma when the first movers cooperate. On the other hand, when first movers defect, the rates of cooperation are 0% among the Korean and U.S. second movers and only 12% among the Japanese.

Suppose that there are two types of players: egoists and reciprocators. While egoists are the selfish maximizers whose motivation and behavior fit the standard game-theoretic assumptions, the reciprocators are intrinsically motivated to reciprocate cooperation with cooperation. The presence of reciprocators changes the game theoretic analysis. Again using backward induction, it can be shown that a first mover is better off by choosing cooperation when the second mover is a known reciprocator.

The problem is, however, how does the first mover know that the second mover is a reciprocator? It would be unreasonable from the evidence available from sequential Prisoner's Dilemma games to assume that everyone is a reciprocator. If the two individuals can talk to each other face-to-face, the second mover will probably try to convince the first mover that he or she would reciprocate if and only if the first mover

cooperates. Can the first mover trust the second mover's promises? Economists use the term "cheaptalk" to describe a first mover's promise and advise the second mover not to trust what is sent. In other words, talking to or looking at others is not viewed as a solution to a social dilemma.

What if the physical symptoms accompanying the promise send reliable signals of the true intentions of those who make the promises? Biological signals, such as facial expressions, body language, eye movement, tone of voice, etc., can be reasonably reliable under certain conditions. At a fundamental level, the signals should be related to the true intention of the signal sender and not under his voluntary (willful) control. Other neurologically based spontaneous signals can serve the purposes as well.

Frank (1988) and, more recently, Schmidt and Cohn (2001) provide the details of such mechanisms. Take facial expression as an example. Several categories of facial muscles are not subject to perfect, conscious control. They do respond to emotions and corresponding neurological processes. Suppose a smile is expected to accompany promises of reciprocation. Those who promise to reciprocate without actually intending to do so may generate smiles, but only by conscious efforts. The muscles used in a conscious effort to generate a smile are different from those muscle that are at work to create genuine smiles. A perceptibly different kind of smile is likely to be produced when an individual is self-consciously trying to smile. Or, if the sender of a smiling signal knows that "artificial smiles" are not good signals and thus is aware that the signal receiver is likely to detect the difference, those who are not intending to reciprocate may not even try to send such artificial signals. If this is the case, a spontaneous smile as a cooperative signal can be quite an effective means of achieving cooperation.

The most observable difference between genuine and artificial smiles is symmetry (see Brown and Moore [2002] for an extensive review of the literature on true and false smiles), but the signal receivers' responses to genuine and artificial smiles are not under complete control of rational thinking. Brown and Moore (2000, 2002) conducted experiments in which symmetric and asymmetric smile icons were shown to experimental subjects. They found that subjects allocate more of their endowed resources to partners to whom symmetric smile icons were associated. The evidence is not however definitive. Eckel and Wilson (2003) report a trust game experiment in which subjects were shown an image of their counterparts projected on a screen. Four different images were shown including either smiling or neutral faces of male or female models. Eckel and Wilson report that even though the subjects revealed in surveys that they trusted smiling faces more than neutral faces, these responses did not correlate with the actual trusting behavior of the subjects.

The theory of mind advocates (Baron-Cohen 1995; Byrne 1995; O'Connell 1998) present a similar account of the ways in which intentions of one person can be revealed to another person. The ability to reason about others' ways of thinking, their intentions, and thus their likely behavior is a distinguishing characteristic of human cognition. A lack of the capacity for mind reading is the key symptom of autism that makes social life almost impossible. Among primates, chimpanzees show some level of mind-reading, compared to monkeys' simple behavior-reading (Cheney and Seyfarth 1990). A substantial difference exists, however, between chimpanzees' level of mind-reading intentionality and that of ordinary adult humans. Using Cheney and Seyfarths's scale, chimpanzees rate at maximum a 1.5 level while humans rate at level 4 (Schmidt and Cohn 2001:188, Box

2). The theory of mind proposes, in addition to intentionality, eye-movements and shared attention as key mind-reading mechanisms.

The existence of individuals who are internally motivated to reciprocate and the potential capacity to detect others' types from physical signals jointly explain the most consistent finding of experimental research on social dilemmas -- that communication enhances cooperation (Ostrom and Walker 1991; Sally 1995). The problem that conditional cooperators face in social dilemmas is the uncertainty regarding whether or not a sufficient number of individuals would also cooperate. This is more than a problem of belief about others' motivations. A conditional cooperator must also be confident that in addition to believing that there are many conditional cooperators in a particular situation, that other conditional cooperators also believe that there are many conditional cooperators, etc.

When people talk to each other and their intrinsic motivations and intentions are reliably revealed to one another, the problem of gaining common knowledge regarding the proportion of reciprocators present can be reduced but not completely eliminated. By making commitments to cooperate, seeing that others also make such commitments, and observing that many of those who make such commitments appear to be trustworthy, a conditional cooperator can be convinced to do his share in a collective endeavor. Reliable signaling serves to facilitate cooperation, along with other cooperation-enhancing functions of communication such as sharing of information about proper ways to cooperate and developing group identity.

Frank et al. (1993) reports a single-shot prisoner's dilemma experiment where subjects were asked to predict their partners' behavior. A group of three subjects were

given thirty minutes to talk to each other. The topic of the talk was not imposed, so the subjects could chat about anything they chose. In all of the taped discussions, everyone made promises that he or she would cooperate in the forthcoming game. After the subjects finished their thirty-minute face-to-face talks, each of them was led to a separate room and asked to predict the likely choices of the other two individuals in a Prisoner's Dilemma game. Frank reports that 73 of the 97 subjects (68%) who were predicted to cooperate actually cooperated in the game. Further, 15 of 25 subjects (60%) who were predicted to defect actually defected. While this capability is obviously not perfect, it is better than chance.

Kikuchi et al. (1997) classify subjects in a one-shot Prisoner's Dilemma into high-, medium-, and low-trusters after administering a pre-experimental survey. They test a hypothesis that high-trusters maintain higher levels of social intelligence and, thus, can more accurately predict the behavior of their partners. In their experiments, groups of six subjects participated in a thirty-minute discussion about garbage collection before they made decisions in single-shot Prisoner's Dilemma games. After the subjects made decisions, they were informed of the identities of their partners and asked to predict the partners' decisions. Kikuchi et al. report that the high-trusters predicted 12 of 16 (75%) cooperators and 10 of 16 (62%) defectors accurately. The accuracy of prediction among the medium- and low-trusters was significantly lower.

Scharlemann et al. (2001) performed a laboratory experiment consisting of a simple two-person, one-shot sequential trust game with monetary payoffs. Each person is shown a photo of his/her partner prior to the game. The photos were chosen from a collection that included those smiling and those not smiling. They find that smiles can increase the

level of trust between strangers significantly, although other facial expressions are also likely to contribute to the cooperative outcomes.

Mealey et al. (1996) find that human subjects have an enhanced memory of faces of cheaters. Black-and-white reproductions of photos of faces of Caucasian males were presented to the subjects together with a fictional descriptive sentence giving information about the depicted individual's status (high or low) and character (related to trustworthiness). One week later they were shown a larger set of photos without descriptions and the subjects tended to recognize non-trustworthy agents more frequently. Oda (1997) conducted a similar study where photos represented partners in one-shot prisoner's dilemma games. One week after the experiment, the subjects were biased to remember those faces that had been portrayed as defectors in the game. DeBruine (2002) performed sequential trust games where the subjects were shown faces of playing partners manipulated to resemble either themselves or an unknown person. Resemblance to the subject's own face raised the expressed trust as a first player in the partners, but had no effect on being a trustworthy or reciprocating second player.

Cosmides (1989) identified biased cognitive processes for identifying cheaters among a cooperating group. Cognitive neuroscience shows that social information is distinct from the processing of other kinds of information, and Stone et al. (2002) describe neurological evidence indicating that social exchange reasoning can be selectively impaired while reasoning about other domains is left intact.

Another recent finding is that the amygdale, which lies within the cerebrum of the brain, is required for accurate social judgment of the facial appearance of others. Individuals with complete bilateral amygdale damage were not able to judge unfamiliar

individuals by visual cues, although this did not hold for verbal descriptions about unknown others (Adolphs et al. 1998). Winston et al. (2002) found a neural basis for trustworthiness judgments using event-related functional magnetic resonance imaging. The neural activities used in trustworthiness judgment may relate to structures that process emotions, although it is not known what cues of facial expressions are important in the process of making trustworthiness judgments (Adolphs 2002).

The existence of reliable signals and their roles in the evolution of intrinsic motives can and have been formalized by economists, generating models of cooperation that are significantly different from Axelrod's earlier evolutionary models. The latter rely on the assumption that the Prisoner's Dilemma is infinitely repeated. As the title of Axelrod's paper -- "The Evolution of Cooperation among *Egoists*" (italics added) – indicates, cooperation does not necessarily require that individuals are motivated by intrinsic preferences. One of the implications of the human ability to detect others' types is that among egoists cooperation cannot really evolve under one-shot or finitely repeated dilemma settings.

An indirect evolutionary approach, foretold by Frank (1987) and fully developed by Werner Güth and his colleagues (Güth and Yaari 1992; Güth 1995; Güth and Kliemt 1998; Güth et al. 2000), provides ways to examine the evolutionary consequences in the presence of the players' ability to detect others' types with more than random accuracy. Güth and Yaari, for example, show that, in the context of a simple sequential trust game, trustworthy types can evolve to be a significant proportion of a population when players can detect others' types with more than random accuracy. Ahn (2002) extends the model to a one-shot sequential Prisoner's Dilemma setting with three preference types.

Populations with egoists and conditional cooperators can be stable under a reasonable parameter range of information and the temptation to defect. Janssen and Stow (nd) show that when simulated agents have the ability to learn to estimate the trustworthiness of others, cooperation with strangers in one-shot games can emerge.

The indirect evolutionary approach is widely utilized by social scientists to provide the logic of viability of the preferences that are different from the rational, selfish preference typically assumed in standard economic modeling. In addition to the evolution of the intrinsic motivation for conditional cooperation, the indirect evolutionary approach can also model costly punishment, observed in a wide range of experiments (Ostrom et al. 1992, 1994; Fehr and Gachter 2002).

SYMBOLS

While signals have immediate, biological attachment to their senders, symbols are secondary, abstract signals constructed socioculturally. The word symbol is derived from the Greek word *symbolon.* In ancient Greece it was a custom to break a slate of burned clay into several pieces and distribute them within a group. When the group reunited the pieces were fitted together (Greek *symbollein*). This confirmed that the individuals were members belonging to the group. Two kinds of symbols are important in facilitating human cooperation: cultural and reputational symbols. We will focus on analyzing the construction of reputational symbols.

Symbols in our usage constitute a subset of signals that are broadly understood in a population, we may call them symbolic signals. Others have used “signs” (Bacharach and Gambetta 2001) or signals (Feldman and March 1981) for the same purpose. Some

socioculturally constructed symbols closely parallel biological signals in that they are immediately observable characteristics of their carriers. They differ from biological signals in that their meanings are socioculturally constructed: tattoos and ties are immediately observable characteristics but they represent different meanings among Hell's Angels and businessmen.

Effective symbolic signals share the characteristics of the effective biological signals. That is, they are hard to fake. Bacharach and Gambetta (2001: 173-174) discuss an incident that one of them experienced at an Oxford college. A group of youngsters outside of the college building claimed that they were to have a seminar in the building, but were locked out. Since the college building hosted many valuable paintings and furniture, the author had to be careful about whether the youngsters' claim was trustworthy. Bacharach and Gambetta proudly report that it took only a split second for the author to assess the trustworthiness of the youngsters' claim using their manifest symbols. The likely symbols – glasses, books, clothes, etc., that Oxford graduate students usually carry – are familiar to an Oxford professor who knew that it would be very costly for an intending group of robbers to coordinate so that all of them manifested such signals. Those signals were effective in that specific sociocultural context in which the senders and receivers of the signals shared a common understanding of the meaning of those signals. The symbols that the professor detected from the group of young people would not have been effective in front of an entrance to a 17th-century Chinese Imperial library.

Reputational symbols are often artificially devised summary information about past behavior and/or other qualifications of a person or a group. Reputational symbols are

essentially information-sharing devices, a solution to the collective action problem that a set of potential transaction partners of an individual or a group face. Viewed as an information-sharing mechanism, reputational symbols represent highly abstract, systemized gossip.

Public documents, such as, for example, the information that Middle Age merchants could obtain from the Law Merchant (Milgrom et al. 1990), are a bridge between gossip and reputational symbols. Various types of public and private documentation that are available to the general public upon request or to a qualified set of individuals, such as visas, drivers' licenses and credit cards, are repositories of information with varying levels of abstraction on the past behavior of certain actors.

Reputational symbols are the most abstract and artifactually condensed information. Let us illustrate the meaning and roles of reputational symbols using the eBay feedback system as an example. The Internet creates a new kind of marketplace with greatly improved information capability that can overcome the limits of conventional markets. The problem of trust among buyers and sellers, however, has become a key obstacle in expanding the scope of online transactions. How can I, as a buyer, be sure that my credit card and other personal information will not be misused? Or, will the seller send me the merchandise at all?

The eBay maintains a "Feedback Profile" that it describes as an "official reputation" for each of its users. The immediate form of the profile is a username/score pair, for example, in the form of "John (125)." This deceptively simple reputation symbol, when reflected upon, reveals how far humans have come to devise systems of

mutually beneficial cooperation, which at the same time protect the system itself and the trustworthy users of it from the potential invasion of untrustworthy exploiters.

First of all, the username, as any name would, assigns symbols to individuals so that each can be identified to any number of others. Animals above a certain evolutionary stage are known to identify others as individuals and store memories specific to each of them. This can be done without language or any cultural devices. Humans, even without taking into account sociocultural mechanisms, have superb capabilities of recognizing the individuality of others through stored memories of appearances and voices. Naming is a step forward. Having each individual named provides further means to share the stored individual-specific information with others who have not had firsthand experience with the named person.

Scores in parentheses in the eBay feedback profile are an even higher-level symbolic representation of the reputational information. They utilize the number system, which by itself is a very modern achievement in the long history of human cultural evolution. The eBay scores are constructed by aggregating comments from the transaction partners of an individual. Comments are coded so that a positive feedback is counted as +1, a negative comment as -1, and a neutral comment as 0. Thus, John (125) is a symbolic condensation of the information that the person using John as the username has received 125 more positive comments than negative ones from the individuals with whom he has done transactions.

There are potential shortcomings of any specific reputation symbols. For example, Malaga (2001) argues that eBay's reputation management is problematic in that it aggregates all positive and negative feedback, leading to overly aggregated information.

There is a barrier to enter for new users as many avoid those sellers with a low reputation score. There is also a potential problem in the accuracy of one's reputation since only half of the participants provide feedback on reputation (Resnick and Zeckhauser 2002). Despite these potential problems, the eBay reputation system works very well in practice. Resnick et al. (nd) show that a high reputation of a seller leads to about 8% higher prices in a controlled experiment with high and low reputation sellers selling the same products.

While eBay-type reputational symbols can be rather simple to construct, other reputational symbols evolve over time by trial and error. Certificates and licenses also serve as reputational symbols that signify, in addition to trustworthiness, the qualification for certain performances. When established and trusted, the certificates and licenses expand the possibilities of mutually beneficial transactions. In the early years of the automobile industry in the United States, the considerable ambiguity due to the lack of standards was an obstacle for the industry's further development. Far more manufacturers existed than there are today, and each firm produced by its own standard. The production quality differed, but consumers could not assess quality before they made the purchases and for some time thereafter. Rao (1994), in his study of the earlier times of the American automobile industry, reports that a series of racing contests organized by national and local newspapers served as a reputation-establishment mechanism as a result of the firms' performance in those contests. Manufacturers' recorded performances, certified by the newspapers as quasi-public institutions, were symbols of the performance of their products and affected the purchase choices of consumers. Eventually, the low performing manufacturer exited from the automobile market and high-performers remained in and developed more reliable standards among them.

The presence of socioculturally constructed reputational symbols, and systems of reputation in general, dramatically alter the behavioral incentives that social actors face in cooperative endeavors. Economists have traditionally assumed that all individuals are engaged in the rational pursuit of their own self-interest and have studied how various reputational mechanism affect behavior of such rational actors (Rubinstein 1979; Kreps et al. 1982; Fudenberg and Maskin 1986; Weigelt and Camerer 1988; Kandori 1992; Tirole 1996).

Reliable reputational symbols, as economists have shown in various ways, induce cooperation by self-interested individuals. One of the implications of these studies is that with reasonably well-functioning reputational symbols, the *behavioral* difference between intrinsically motivated and extrinsically motivated types may disappear: with reliable symbols, self-interested individuals will cooperate out of their own selfish concerns, i.e., to reap the long-term benefit of sustained cooperation. Reliable symbols further assure the intrinsically motivated conditional cooperators that a large proportion of others will also cooperate. In terms of behavior and relative payoffs, therefore, the logical conclusion is that different types of players behave in a similar manner and obtain the same payoffs in the presence of a reliable reputation system.

If we consider the possibility that many humans do not possess the level of prudence needed to behave rationally, a great deal of which involves resisting immediate gratification to assure future gains, it is not too unreasonable to hypothesize that the reputation systems do have different effects on different types. That is, while the intrinsically motivated cooperators may find it easier to resist the temptation to defect, the level of self-restraint that self-interested individuals need may be significantly higher.

Devising and maintaining reliable systems of reputational symbols involve collective action problems among many individuals. The eBay reputation profile depends on the users' taking time to provide accurate inputs on their transaction partners' trustworthiness. Symbols are in that sense a public good, like all language, which once provided benefits to everyone whether or not they contributed their time and resources for the establishment of the reputation symbols.

Symbols evolve for diverse reasons. Over time, people learn to mimic the symbols and, thus, nullify their effectiveness. As symbols lose their effectiveness, the incentives to behave in a trustworthy manner decrease. Thus, the trustworthy users of a system of symbols find it necessary to constantly evolve the symbolic systems of reputation. The evolutionary arms races that are observed in the biological world also occur in the sociocultural world. And the quality of any reputational symbol as a public good may erode over time if substantial investment is not made in "policing" the use of the system. Recent events in American corporate industry (Enron et al.) have called the reputational symbols awarded by auditing firms into serious question.

CONCLUSION

The great potential of human cooperation is rooted in the evolved human biological capacity to use signals. This potential that is the foundation for all society is further developed by socioculturally constructed reputational symbols. Both signals and symbols are not just devices for cooperation but essentially serve as the mechanisms by which intrinsically motivated conditional cooperators can evolve to compose a large

proportion of a population. At the same time, neither biological signals nor sociocultural symbols work perfectly. They are both prone to errors and abuses.

The scope and scale at which a society can maintain cooperative endeavors greatly affect the society's destiny. In a world where the types of social interactions change so rapidly, it is important to craft sophisticated systems of rules and symbols so that society will not enter a Hobbesian world in which predation replaces cooperation. A world that becomes highly affected by the internet, terrorist networks and a global market for commodities faces immense challenges in developing reliable symbolic systems that allow trustworthy actors to detect each other to maintain human sociality.

ACKNOWLEDGMENTS

We gratefully acknowledge support from the Center for the Study of Institutions, Population, and Environmental Change at Indiana University through National Science Foundation grants SBR9521918 and SES0083511.

REFERENCES CITED

Adolphs R. 2002. Trust in the brain. Nature neuroscience, 5(3): 192-193.

Adolphs R, Tranel D, and Damasio A. 1998. The human amygdala in social judgement. Nature 393: 470-474.

Ahn TK. 2002. Information and the Evolution of Preferences in One-shot Prisoner's Dilemma. Working Paper. Bloomington: Indiana University, Workshop in Political Theory and Policy Analysis.

Ahn TK, Ostrom E, and Walker J. In press. Incorporating Motivational Heterogeneity into Game-theoretic Models of Collective Action. Public Choice.

Andreoni J, and Miller JH. 1993. Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence. Economic Journal 103: 570-585.

Axelrod R. 1981. The Emergence of Cooperation among Egoists. American Political Science Revie*w* 75: 306-318.

Axelrod R, and Hamilton WD. 1981. The Evolution of Cooperation. Science 211:1390-1396.

Bacharach M, and Gambetta D. 2001. Trust in Signs. In: Cook C, editor. Trust in Society. New York. Russell Sage Foundation. pp. 148-184.

Baron-Cohen S. 1995. Mindblindness: An Essay on Autism and Theory of Mind. Boston: MIT Press.

Bendor J, and Swistak P. 1997. The Evolutionary Stability of Cooperation. American Political Science Review 91:290-307.

Boyd R, and Lorberbaum J. 1987. No Pure Strategy Is Evolutionarily Stable in the Repeated Prisoner's Dilemma Game. Nature 327:58-59.

Brown WM, and Moore C. 2000. Is Prospective Altruist-detection an Evolved Solution to the Adaptive Problem of Subtle Cheating in Cooperative Ventures? Supportive Evidence Using the Wason Selection Task. Evolution and Human Behavior 21:25-37.

Brown WM, and Moore C. 2002. Smile Asymmetry and Reputation as Reliable Indicators of Likelihood to Cooperate: An Evolutionary Analysis. Advances in Psychology Research 11:59-78.

Byrne RB. 1995. Thinking Primates. Oxford: Oxford University Press.

Cheney DL, and Seyfarth RM. 1990. How Monkeys See the World. Chicago: Chicago University Press.

Clutton-Brock T. 2002. Breeding Together: Kin Selection and Mutualism in Cooperative Vertebrates. Science 296: 69-72.

Cosmides L. 1989. The logic of social exchange: Has selection shaped how humans reason? Studies with the Wason selection task. Cognition 31: 187-276.

Darwin C. 1873 [1872]. The Expressions of the Emotions in Man and Animals. New York: D. Appleton.

DeBruine LM. 2002. Facial resemblance enhances trust. Proceedings of the Royal Society London B 269: 1307-1312.

de Waal FBM. 1989. Food Sharing and Reciprocal Obligations among Chimpanzees. Journal of Human Evolution 18: 433-459.

de Waal FBM. 1997. Bonobo: The Forgotten Ape. University of California Press. Berkeley, CA.

Eckel CC, and Wilson RK. 2003. The Human Face of Game Theory: Trust and Reciprocity in Sequential Games. In: Ostrom E, and Walker J, editors. Trust,

Reciprocity, and Gains from Association: Interdisciplinary Lessons from Experimental Research, Chapter 9. New York: Russell Sage Foundation.

Fehr E. and Gächter S. 2002. Altruistic punishment in humans. Nature 415: 137-140.

Feldman MS, and March JG. 1981. Information in Organizations as Signals and Symbols. Administrative Science Quarterly 26:171-186.

Frank RH. 1987. If Homo Economicus Could Choose His Own Utility Function, Would He Want One with a Conscience? American Economic Review 77(4):593-604.

Frank RH. 1988. Passions within Reason: The Strategic Role of the Emotions. New York: Norton.

Frank RH, Gilovich T, and Regan D. 1993. The Evolution of One-Shot Cooperation: An Experiment. Ethology and Siociobiology 14: 247-256.

Fudenberg D, and Maskin E. 1986. The Folk Theorem in Repeated Games with Discounting or with Incomplete Information. Econometrica 54: 533-554.

Güth W. 1995. An Evolutionary Approach to Explaining Cooperative Behavior by Reciprocal Incentives. International Journal of Game Theory 24:323-344.

Güth W, and Yaari M. 1992. An Evolutionary Approach to Explaining Reciprocal Behaviour in a Simple Strategic Game. In: Kliemt, H. editor, Explaining Process and Change. Ann Arbor: University of Michigan Press. pp. 23-34

Güth W, and Kliemt H. 1998. The Indirect Evolutionary Approach: Bridging the Gap Between Rationality and Adaptation. Rationality and Society 10(3):377-399.

Güth W, Kliemt H, and Peleg B. 2000. Co-evolution of Preferences and Information in Simple Games of Trust. German Economic Review 1(1):83-110.

Henrich J, Boyd R, Bowles S, Camerer C, Fehr E, Gintis H, McElreath R, Alvard M, Barr A, Ensminger J, Hill K, Gil-White F, Gurven M, Marlowe F, Patton JQ, Smith N, and Tracer D. 2001. Economic Man in Cross-Cultural Perspective: Behavioral Experiments in Fifteen Small-Scale Societies. Santa Fe Institute Working paper 063.

Isaac RM, and Walker JM. 1988. Group Size Effects in Public Goods Provision: The Voluntary Contribution Mechanism. Quarterly Journal of Economics 103:179-200.

Janssen MA. and Stow DW. nd. Evolution of Cooperation in a One-Shot Prisoner's Dilemma Based on Recognition of Trustworthy and Untrustworthy Agents. submitted

Kandori M. 1992. Social Norms and Community Enforcement. Review of Economic Studies 59: 63-80.

Kikuchi M, Watanabe Y, and Yamagishi T. 1997. Judgment Accuracy of Others' Trustworthiness and General Trust: An Experimental Study. (in Japanese with an English abstract). Japanese Journal of Experimental Social Psychology 37:23-36.

Kreps DM, Milgrom P, Roberts J, and Wilson R. 1982. Rational Cooperation in the Finitely Repeated Prisoner's Dilemma. Journal of Economic Theory 27: 245-252.

Kurzban R. 2003. Biological Foundations of Reciprocity. In: Ostrom E, and Walker J, editors. Trust, Reciprocity, and Gains from Association: Interdisciplinary Lessons from Experimental Research. pp. 105-127. New York: Russell Sage Foundation.

Lorberbaum J. 1994. No Strategy Is Evolutionarily Stable in the Repeated Prisoner's Dilemma. Journal of Theoretical Biology 168(May):117-130.

Malaga RA. 2001. Web-based reputation management systems: problems and suggested solutions. Electronic Commerce Research 1: 403-417.

Mealey L, Daood C, and Krage M. 1996. Enhanced memory for faces of cheaters, Ethology and Sociobiology 17: 119-128.

Milgrom P, North D, and Wengast B. 1990. The role of institutions in the revival of trade: the law merchant, private judges, and the champaign fairs. Economics and Politcs 2: 1-23.

O'Connell S. 1998. Mindreading: An Investigation into How We Learn to Love and Lie. New York: Doubleday.

Oda R. 1997. Biased Face Recognition in the prisoner's Dilemma, Evolution and Human Behavior, 18: 309-315.

Ostrom E., Gardner R, and Walker J. 1994. Rules, Games, & Common-Pool Resources. Ann Arbor: The University of Michigan Press.

Ostrom E, and Walker J. 1991. Communication in a Commons: Cooperation without External Enforcement. In: Palfrey (editor). Laboratory Research in Political Economy, pp. 287-322. Ann Arbor: University of Michigan Press.

Ostrom E., Walker J, and Gardner R. 1992. Covenants With and Without a Sword: Self-Governance is Possible. American Political Science Review 86(2): 404-417.

Rao H. 1994. The Social Construction of Reputation: Cerfitication Contests, Legitimation, and The Survival of Organizations in the American Automobile Industry: 1895-1912. Strategic Management Journal 15: 29-44.

Resnick P, and Zeckhauser R. 2002. Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System. In: Baye MR (editor) The Economics of the Internet and E-Commerce, pp. 127-157. Amsterdam: Elsevier Science.

Resnick P, Zeckhauser R, Swanson J, and Lockwood K. nd. The value of reputation on eBay: A controlled experiment. Unpublished manuscript (2002) University of Michigan. http://www.si.umich.edu/~presnick/papers/postcards/

Rilling JK, Gutman DA, Zeh TR, Pagnoni G, Berns GS, and Kilts CD. 2002. A Neural Basis for Social Cooperation. Neuron 35: 395-405.

Rubinstein A. 1979. Equilibrium in Supergames with the Overtaking Criterion. Journal of Economic Theory 21: 1-9.

Sally D. 1995. Conversation and cooperation in social dilemmas: A meta-analysis of experiments from 1958 to 1992. Rationality and Society 7: 58-92.

Scharlemann JPW, Eckel CC, Kacelnik A, and Wilson RK. 2001. The value of a smile: Game theory with a human face. Journal of Economic Psychology 22: 617-640.

Schmidt D, Shupp R, Walker J, Ahn TK, and Ostrom E. 2001. Dilemma Games: Game Parameters and Matching Protocols. Journal of Economic Behavior and Organization 46:357-377.

Schmidt KL, and Cohn JF. 2001. Human Facial Expressions as Adaptations: Evolutionary Questions in Facial Expression Research. Yearbook of Physical Anthropology 44:3-24.

Selten R, and Stoecker R. 1986. End Behavior in Sequences of Finite Prisoner's Dilemma Supergames: A Learning Theory Approach. Journal of Economic Behavior and Organization 7: 47-70.

Stephens DW, McLinn CM, and Stevens JR. 2002. Discounting and Reciprocity in an Iterated Prisoner’s Dilemma. Science 298: 2216-2218.

Stone VE, Cosmides L, Tooby J, Kroll N, and Knight RT. 2002. Selective impairment of reasoning about social exchange in a patient with bilateral limbic system damage. Proceedings of the National Academy of Science 99: 11531-11536.

Tirole J. 1996. A Theory of Collective Reputations (with Applications to the Persistence of Corruption and to Firm Quality). Review of Economic Studies 63(1): 1-22.

Trivers RL. 1971. The Evolution of Reciprocal Altruism. The Quarterly Review of Biology 45(4):35-57.

Weigelt K, and Camerer C. 1988. Reputation and Corporate Strategy: A Review of Recent Theory and Applications. Strategic Management Journal 9(5): 443-454.

Wilson EO. 2000 [1975]. Sociobiology: the new synthesis. Harvard University Press, Cambridge, MA.

Winston JS, Strange BA, O'Doherty J, and Dolan RJ. 2002. Automatic and intentional brain responses during evaluation of trustworthiness of faces. Nature neuroscience 5(3): 277-283.