

# 17

## Renaissance of the Individual *Reciprocity, Positive Assortment, and the Puzzle of Human Cooperation*

DOMINIC D. P. JOHNSON,  
MICHAEL E. PRICE, AND  
MASANORI TAKEZAWA

At the end of all our journeying, we will arrive at the original place and know it for the first time.

—T. S. Eliot

When recognized, adaptation should be attributed to no higher a level of organization than is demanded by the evidence.

—G. C. Williams

Cooperation determines the success or failure of human interactions ranging from everyday social exchange, to global trade, environmental management, and international conflict. Yet, strangely enough, this fundamental aspect of human behavior remains poorly understood and a topic of great debate spanning several disciplines. Since the mid-1960s, cooperation was widely understood as promoting the inclusive fitness of the individual cooperator. However, in the past few years, researchers have asserted that the unprecedented extent of cooperation among human groups of nonkin (e.g., in collective actions) lies beyond the explanatory power of individual-level theories. These researchers argue that this unusual level of cooperation demands a special, higher level explanation such as

biological or cultural group selection. While humans are indeed unusually cooperative, we outline evidence suggesting that cooperation is best explained not in terms of group-level adaptation, but rather by applying, extending, and synthesizing existing theories of individual-level adaptation. Specifically, we contend that (1) most humans are “conditional reciprocators,” who cooperate to whatever extent their interaction partners do (a Triversian *reciprocity* model) and (2) cooperators can achieve “positive assortment” and, thus, outcompete free riders, by preferentially interacting with other cooperators (a Hamiltonian *positive assortment* model). Previous models have tended to examine Triversian and Hamiltonian processes in isolation, yet the synergistic effects between them may account for the evolution and maintenance of collective action. In this chapter, we present (1) an overview of evolutionary accounts of human cooperation and how they have changed over time, (2) a *theoretical* model demonstrating how a strategy for continuous (as opposed to binary) reciprocity in sizable groups could replicate successfully in a population of competing strategies, and (3) *empirical* evidence from several different disciplines which suggests that most people do in fact act as positively assorting reciprocators in group interactions. Group selection is possible in theory but remains unnecessary to explain the diversity and prevalence of human cooperation. As we argue, the current focus on group selection stems from an erroneous fork along the road of research on human cooperation, traceable to a specific influential model of reciprocity by Robert Boyd and Pete Richerson in 1988. Our initial exploration of the alternative path suggests there are fruitful avenues for a much improved understanding of human cooperation.

### DARWIN'S PUZZLE

Charles Darwin worried that his entire theory of evolution by natural selection might be blown out of the water by a phenomenon apparently prevalent in nature: altruism. If evolution proceeded through the survival of the fittest individuals, then why would bees, for example, evolve to engage in suicidal stinging behavior in defense of their colonies? Biologists after Darwin did not worry about this too much because they tended to see evolution as working for “the good of the species” or “the good of the group,” a self-regulating machine in which altruism made perfect sense in the continuation of a greater cause. Only later was this perception revealed to be fatally flawed. As biologists stripped away the veiled mechanisms of DNA and genetic replication and recombination, whose effects Darwin predicted without knowing the cause, it became clear that traits that benefited the species as a whole would quickly disappear from the gene pool. Since genetic information is passed directly only from parents to offspring, any strategy so foolish as to sacrifice itself for unrelated others will die a quick death. Rejection of the good of the group argument, however, brought biologists back to the original puzzle: how can altruism evolve? Sadly for Darwin, this problem, which vexed him so much, was not solved until long after his death. But since the 1960s, the foundation of the gene as the unit of selection, and the predominance of individual-level adaptationism paved the way for a new understanding of altruism/cooperation via four specific theories described in the following section: kin selection (and genic self-favoritism in general), reciprocal altruism, indirect reciprocity, and costly signaling.

#### *Kin Selection (and Genic Self-Favoritism in General)*

In 1964, W. D. Hamilton published his famous insight that cooperation could evolve via kin selection. Since a gene for altruism could spread by promoting the replication of copies of itself contained in other individuals, and because close genetic kin will be relatively likely to carry copies of the same gene for altruism, promoting the survival and reproduction of one's relatives promotes the survival and reproduction of one's own genes (for more on kinship and social behavior, see West et al., this volume). This simple idea went on to revolutionize biology and anthropology (Chagnon & Irons, 1979; E. O. Wilson, 1975, 2000). Hamilton's insight is more profound than is often realized, because his theory of kin selection is actually just one (albeit very important) application

of his more general theory of “genetic self-favoritism.” The more general theory can in principle explain altruism not only among close genetic kin, but also among nonkin, as long as they (1) share the same gene(s) for altruism and (2) are able to interact preferentially with each other. Dawkins’ (1976) famous example of genetic self-favoritism among nonkin imagined a gene that produced a recognizable phenotypic label (a green beard) in its bearers, as well as a motivation to cooperate with others who have green beards. Although regarded by many as implausible during the time at which Hamilton’s ideas were being disseminated, apparent examples of green-beard altruism have recently surfaced in the literature (Haig, 1996; Keller, 2002; Keller & Ross, 1998; Queller, Ponte, Bozzaro, & Strassmann, 2003; Sinervo et al., 2006).

### *Reciprocal Altruism*

A few years after Hamilton published his theory, Robert Trivers (1971) showed that cooperation could also evolve if a cooperative act by a donor toward a beneficiary were reciprocated by the beneficiary at a later time. Such mutually beneficial reciprocal altruism, in theory, could evolve in the complete absence of the genetic self-favoritism described by Hamilton’s kin selection—indeed, even between different species. Trivers’ ideas were later supported in formal modeling by Axelrod and Hamilton (1981) and, like kin selection, dramatically altered the research agendas and understanding of animal and human interactions in biology, anthropology, and the social sciences (see Trivers, 2006).

### *Indirect Reciprocity*

Richard Alexander (1987) later extended the logic of Trivers’ reciprocal altruism by noting that altruistic acts may advertise a propensity to cooperate and, via a process of “indirect reciprocity,” attract future cooperation *from third parties*. Third parties may acquire information about an individual’s cooperativeness either via direct observation of a reciprocal exchange, or by acquiring reputational information about an individual’s past history of cooperative behavior. Like Trivers’ theory, Alexander’s ideas eventually found support from formal models (Nowak & Sigmund, 1998; Panchanathan & Boyd, 2004), as well as experimental evidence that people cooperate more when their interaction history is publicly advertised (Nowak & Sigmund, 2005; Wedekind & Milinski, 2000).

### *Costly Signaling*

Alexander (1987) and Trivers (1971) both focused on what kinds of benefits might be supplied and acquired by altruists, and both identified cooperative acts themselves as the primary benefit: Altruistic donors receive more altruistic donations, via direct and indirect reciprocity, because others prefer them as cooperative partners. Amotz Zahavi (1975, 1995), however, pointed out that the return benefit acquired by altruists need not be altruism per se. Altruism could be a “costly signal” that is used to advertise an attractive genetic quality that is only incidentally related to altruism itself (the logic being that only most fit individuals can afford to help others on top of the demands of their own survival). Such signaling may be rewarded by forms of fitness benefit other than altruism, such as increased mating opportunities (Gintis, Smith, & Bowles, 2001; Zahavi, 1995).

Kin selection (and to a lesser extent genetic self-favoritism in general), reciprocal altruism, indirect reciprocity and costly signaling have since routinely been used to explain formerly puzzling examples of animal cooperation in a wide diversity of nonhuman species and contexts (for a review, see Dugatkin, 1997; Gadagkar, 2001; Wilson, 2000). More than a century after Darwin, the first puzzle of cooperation has largely been solved.

## THE NEW PUZZLE OF COOPERATION

By contrast with cooperation among nonhuman animals, cooperation *among humans* is still relatively poorly understood. Although people do increase cooperation when the four “traditional” theories of cooperation—kin selection, reciprocal altruism, indirect reciprocity, and costly

signaling—predict that we should do so, we nevertheless continue to cooperate even when these theories would seem to predict otherwise (Fehr & Fischbacher, 2003; Gintis, 2003). The key evidence for this puzzling behavior comes from controlled experimental studies demonstrating that subjects cooperate even when experimenters have tried to carefully eliminate any possible self-interested incentives to cooperate. For example, even in games in which reputational gains are impossible because all players are anonymous and in which reciprocity is impossible because all interactions are one-shot, subjects continue to behave cooperatively to some extent—not as much as they do when reputation and reciprocity are permitted, but to a significant extent nevertheless. In the words of Fehr and Gächter (2002), “[P]eople frequently cooperate with genetically unrelated strangers, often in large groups, with people they will never meet again, and when reputation gains are small or absent,” leaving human cooperation as an “evolutionary puzzle” (p. 137). A growing experimental literature agrees that, when asked to play simple games that are designed to represent everyday social dilemmas, people from both modern and preindustrial societies around the globe cooperate to a greater extent than can be accounted for by traditional theories of cooperation—a phenomenon dubbed “strong reciprocity” (Fehr & Fischbacher, 2003; Gintis, 2000; Henrich et al., 2004). So far, no one has come up with a consensus explanation for this apparent excess of cooperation.

In recent years, two camps have arisen that explain human cooperation in different ways. Both camps agree on the empirical evidence that, in laboratory situations, humans exhibit a propensity to engage in voluntary and costly cooperation and punishment of defectors. The debate is not *if* people do this but *why*. The first camp, the *group*-level adaptationist camp, believes that the traditional theories of cooperation—which are based on individual-level adaptation—*cannot* explain human cooperation. From this, they deduce that human cooperation must serve group purposes. They therefore resort to some form of biological or cultural group selection as a possible explanation (Boyd, Gintis, Bowles, & Richerson, 2003; Fehr & Fischbacher, 2003; Gintis, 2000). The second camp, the *individual*-level adaptationist camp, believes that the traditional theories based on individual-level adaptation *can* explain human cooperation. They maintain that the experimental economic evidence is being interpreted wrongly and that, with an alternative perspective, individual-level theories can explain human cooperation after all (Burnham & Johnson, 2005; Hagen & Hammerstein, 2006; Johnson, Stopka, & Knights, 2003; Price, Cosmides, & Tooby, 2002).

The divergence between the group- and individual-level camps is so significant for theoretical understanding, as well as practical applications, that it is essential to try and work through the arguments and resolve the differences. In doing so, we must differentiate the two general ways in which the camps disagree: First, they disagree about the interpretation of the experimental data; second, they disagree about the theoretical grounds for suggesting that human cooperation evolved via individual-level adaptation. These two aspects of disagreement are discussed in turn in the following section.

### *First Disagreement: Do the Experimental Data Refute Individual-Level Adaptationist Theories of Human Cooperation?*

As previously noted, subjects in experimental games continue to engage in costly cooperation and punishment even when game designers have attempted to eliminate incentives that should elicit such behavior (according to individual-level adaptationist theories). Group-level adaptationists regard these results as evidence that cooperative behavior must be generated by adaptations that evolved to serve the interests of biological or cultural groups. From the individual-level perspective, however, a big problem remains: Even if subjects consciously understand and believe that there are no personal consequences of their actions—for example, that the game is anonymous and they can do what they like with impunity the semiautonomous psychological mechanisms that regulate their cooperative behavior may remain alert and sensitive to cues of social exchange. Subjects cannot simply *switch off* the psychological mechanisms that guide our behavior in everyday social exchanges—they cannot leave their brains at the laboratory door (Trivers, 2004, 2006; Vogel, 2004). The experimental

evidence, in fact, does suggest that subliminal cues of social exchange—for example, the presence of stylized eye spots on the wallpaper of computer screens or of a robotic head with human-like eyes in experimental laboratories—do increase levels of cooperation (Bateson, Nettle, & Roberts, 2006; Burnham & Hare, in press; Burnham, 2003; Haley & Fessler, 2005).

A major problem with the group-level adaptationist interpretation is that it confuses ultimate and proximate causes of behavior (Tinbergen, 1968; see Table 17.1). They expect that if psychological mechanisms evolved as individual-level adaptations, then these mechanisms should function adaptively as such in every environment (focusing exclusively on the “ultimate” explanation). However, they overlook the fact that these semiautonomous mechanisms are capable only of responding to informational cues that would have been present in the environments to which they are adapted (which redirects our attention toward “proximate” explanations). Our psychology simply fails to optimize behavior in evolutionarily novel circumstances (e.g., laboratory experiments or big cities), and better reflects the constraints of the environments in which our psychological mechanisms for cooperation evolved, environments characterized by small groups of extended kin, few strangers, strong hierarchies, and lasting reputations (Barkow, Cosmides, & Tooby, 1992; T. Burnham & Johnson, 2005; Hagen & Hammerstein, 2006; Johnson et al., 2003). Of course, thinking human beings are perfectly able to *adjust* their cooperation to situations in which these social contexts are at stake, but if even part of our response is subconscious, then we continue to behave *as if* they are at stake even where they are not.

A further problem with the group-level adaptationist interpretation revolves around the significance and ecological validity of one-shot encounters, which are often central to laboratory demonstrations of costly cooperation behavior (Hagen & Hammerstein, 2006). Group-level adaptationists like to argue that one-shot interactions were common or important in our evolutionary history, because this would imply that humans have adapted to deal with them. If so, there is not necessarily any mismatch between cooperative behavior in one-shot experimental games and our evolved psychological mechanisms. This would, they argue, rule out the possibility that people “mistake” one-shot experimental encounters for repeated encounters (behaving *as if* in an ongoing reciprocal exchange). Instead, they should recognize a one-shot encounter for what it is and withhold cooperation. This line of reasoning has spawned a considerable debate over how common such one-shot encounters were in human evolutionary history.

**Table 17.1 Tinbergen’s Four Causes of Behavior**

Cause	Description	Relevance for Cooperation?	Possible Tests
Ultimate	The behavior’s evolutionary function (e.g., building nests to protect eggs)	What is the adaptive significance of cooperation?	Experiments that identify the contexts and payoffs associated with cooperation
Proximate	The behavior’s immediate trigger (e.g., hormones)	What are the psychological motives behind cooperation?	Experiments that elicit subconscious triggers of cooperation
Ontological	The behavior’s developmental or learned components (e.g., observing or imitating parents)	How does childhood development and learning influence cooperation?	Experiments with children of different ages; comparative studies of communities with different teaching methods
Phylogenetic	The behavior’s distribution among closely related species and common ancestors	What are the differences and similarities between human and primate/other animals’ cooperation?	Cross-species comparisons of cooperation behavior

However, this argument has run into a number of problems. Firstly, encountering a stranger does not necessarily represent a one-off encounter. A single meeting may (1) actually *cause* an increased probability of positive or negative encounters in the future (Hagen & Hammerstein, 2006) or (2) *contain* multiple mini-reciprocal exchanges operating over seconds or minutes, which could make the difference between offering a banana to make a new friend or threatening violence and getting killed (Trivers, 2004). Secondly, being told that games are a one-shot encounter may fail to cause the games to be recognized as such by experimental subjects, given that they are usually consciously or unconsciously aware that they are playing with fellow students or villagers (Hagen & Hammerstein, 2006). Finally, a hunter-gatherer's one-shot encounter implies by definition an interaction with a *nongroup* member and, thus, would hardly be expected to elicit prosocial behaviors that are held to evolve for *within-group* cooperation in competition with other groups (Hagen & Hammerstein, 2006). At best, the significance of one-shot encounters can be interpreted both ways. At worst, it is a meaningless distraction from the central issue, especially given that we will never know what the frequency of one-shot encounters really was in our evolutionary past.

Regardless of how rare or common such circumstances may have been in our ancestors' lives, it is reasonable to doubt that these circumstances are realistically recreated in experimental laboratories that are crowded with other people (who, anonymous or not, visible or not, are nevertheless present), who are playing some sort of interaction game *with you* (and, therefore, have an explicit interest in your actions), and in which your behavior is being recorded by computers and scrutinized by scientists. The techniques that experimenters use to create an impression of anonymity (e.g., telling subjects that the experiment is anonymous, or implementing a double-blind methodology) may not represent valid cues of anonymity to subjects' semiautonomous psychological mechanisms for cooperation.

### *Second Disagreement: Could Selection Have Favored Individual-Level Adaptations for Cooperation in Humans?*

The different interpretations of the experimental data just described are influenced by different theoretical orientations: Only the individual-level adaptationists, and not the group-level adaptationists, believe that individual-level adaptations were predominant in the natural selection of cooperation behavior in humans. To understand the source of these differences of opinion, it is first necessary to consider what is special about human cooperation. Both the individual- and group-level adaptationists agree that to an extent that is apparently unique among animals, humans cooperate in large groups of unrelated individuals to produce shared resources (hereafter, "collective actions"). In considering why many scientists consider human cooperation to be an evolutionary puzzle, it is important to remember that it is this *specific kind* of interaction—collective actions—that they consider puzzling. Simpler kinds of cooperative interactions, for example dyadic exchange, are relatively easy to explain via conventional models of direct or indirect reciprocity (Axelrod & Hamilton, 1981; Nowak & Sigmund, 1998; Panchanathan & Boyd, 2004).

The special structure of collective actions requires special models to explain how cooperators avoid being exploited to extinction by free riders. Formal modelers are yet to demonstrate that the individual-level adaptive strategies that allow cooperation to evolve in simple dyadic contexts, such as direct and indirect reciprocal strategies, can promote the evolution of cooperation in collective actions (although the potential to do so is great; Nowak & Sigmund, 2005). Further, one very influential model (Boyd & Richerson, 1988) suggests explicitly that collective action *cannot* evolve via reciprocity. On the other hand, group-level adaptationists have produced several models demonstrating that collective action *could* evolve via biological or cultural group selection (Boyd et al., 2003; Gintis, 2000, 2003). At this point, therefore, the group-level adaptationists seem to wield a theoretical advantage over the individual-level adaptationists in the extent to which they have expressed their theories formally.

That being the case, it remains essential to explore individual-level theories to the full (and this has simply not been done). If we trace the history of the literature on human cooperation, we can see that some wrong turns were made in the path that led to the current popularity of group selection. By exploring the untrodden paths, we find that cooperation may in fact be explicable in terms of individual-level theories, with a high degree of parsimony with experimental data. There are three primary reasons why such a “renaissance of the individual” is likely to improve our comprehension of human cooperation: (1) Group selection models are generally not parsimonious as biological adaptationist explanations, (2) existing individual-level theories can be extended and synthesized to produce successful formal models of collective action, and (3) behavioral data from real-life collective actions are better predicted by individual-level rather than group-level models. Each of these arguments is developed in detail in the following sections.

### GROUP SELECTION VERSUS PARSIMONY

Once ridiculed as a wrong-headed misunderstanding of how selection works (Williams, 1966, 1992), group selection has enjoyed a resurgence of interest since the 1980s (Wilson, 2001; Wilson & Sober, 1994), and has attracted particular attention in studies of human cooperation (Boyd et al., 2003; Gintis, 2000; Sober & Wilson, 1998; Wilson, 2002). Three things have changed to make this possible. First, there was a realization that, in purely theoretical terms, group selection might work after all (Wilson, in press). Second, there was a realization that the selection of *cultural* traits, rather than genes alone, may be driven by group selection (Boyd & Richerson, 1985; Richerson & Boyd, 2004). The evolution of cultural traits was, of course, seen as particularly relevant to our own culturally rich species, where innovations in one group can lead to advantages over other groups. The third change was the general realization that group selection need not compete directly with the often opposing selective forces working at the individual level (Sober & Wilson, 1998). Rather a process of multilevel selection is envisaged, in which group traits may come to the fore at certain times. This new work on group selection raises the opportunity to explain a number of otherwise puzzling traits by invoking their benefit in terms of the good of the group.

Darwin himself mused on the possibility of group selection (partly for want of an alternative explanation for altruism). Without any knowledge of DNA, the idea seemed perfectly plausible and conformed to the logic of natural selection. Especially with respect to humans, he saw that, in principle, human groups that shared certain traits that favored the in-group may spread at the expense of more selfish groups who were less able to work together. This, he thought, might represent an evolutionary origin for the development of human morality (Darwin, 1871).

It is crucial to remember, however, that we need only invoke mechanisms that are sufficient to explain the puzzle at hand. Group selection may or may not explain human cooperation. But before it is invoked, are we certain that individual-level adaptationism cannot explain human cooperation? The principle of Occam’s razor advises us to be sure that the simpler explanations can be ruled out before going on to more complex and controversial explanations that contain more assumptions (as group selection does). George Williams (1966) took this parsimonious approach when he stated, in the opening words of *Adaptation and Natural Selection*, “[W]hen recognized, adaptation should be attributed to no higher a level of organization than is demanded by the evidence” (p. v). In other words, our evolutionary accounts will more likely be correct if we thoroughly attempt to explain adaptation in terms of lower level selective processes, before we resort to higher level explanations.

### BACK TO SQUARE ONE: AN INDIVIDUAL-LEVEL ADAPTATIONIST EXPLANATION OF COOPERATION IN GROUPS

As previously noted, cooperation between two individuals is relatively easy to explain via reciprocity, the real problem is explaining the evolution of *n*-player collective actions. In such groups,

cooperation can quickly break down because while there may be an attractive reward if everyone cooperates, cheats can do better than cooperators by withholding their contributions while nevertheless benefiting from the public good (Olson, 1965; Ostrom, 1990). For example, a cheat can fail to pay taxes and still drive on public roads. More formally, in a group of  $n$  members, one's cost of contributing  $c$  creates a public good, with a total benefit  $mc$  that is shared equally by all members (where  $m$  is some positive number). One can contribute productively when  $m > 1$ ; however, if  $n > m > 1$ , then one can profit more individually by contributing nothing, and free riding on the contributions of comembers.

How can cooperators in collective actions harvest the benefits of their production, without allowing free riders to exploit them to extinction? The solution to this problem, in general, is to cooperate preferentially with other cooperators instead of with free riders and, thus, to exclude free riders from the benefits of production. This principle of preferential cooperation among cooperators is a basic law of adaptive cooperation, and it underlies the evolutionary success of Trivers' (1971) reciprocal altruism, which allows cooperators to sustain interactions only with other cooperators while avoiding benefiting free riders. This principle also explains the adaptiveness of Alexander's (1987) indirect reciprocity and Hamilton's (1964) genic self-favoritism among nonkin, both of which allow cooperators to preferentially assort with other cooperators and exclude free riders from the circle of mutual cooperation, that is, to engage in *positive assortment*.

Given the adaptive potential of reciprocity and positive assortment, why have they been regarded by many researchers as inadequate explanations for collective action? In the case of positive assortment, the argument is not about whether systems requiring cooperators to identify and assort with each other could in principle allow collective action to evolve; rather, it is about whether such systems could (1) survive invasions by deceptive individuals posing as cooperators (the *deception problem*) and (2) involve a mechanism for efficiently ostracizing free riders who wanted to remain in a group of cooperators (the *ostracization problem*).

The deception problem arises because in a system of positive assortment, cooperators must have some means of recognizing other cooperators, in order to interact with them preferentially. Cooperators, therefore, must display some phenotypic label that identifies them as cooperators. If this label can be easily faked and displayed by noncooperators, then the system will be vulnerable to deception: Mutations that produce the label of cooperativeness but do not produce the cooperative behavior itself, will acquire the benefits of cooperation without paying the costs and, thus, will exploit honest label displayers to extinction. The deception problem will be more serious in a system involving a label that is more arbitrary and less intrinsically indicative of cooperative behavior itself, e.g., Dawkins' (1976) "green beard". The solution to this problem would be for cooperators to rely on a label that is a maximally nonarbitrary and difficult-to-fake signal of cooperativeness, ideally convincing displays of altruistic behavior itself (Dawkins, 1976; Price, 2006).

The seriousness of the ostracization problem in a system of positive assortment will depend heavily on the scale and type of collective action. For instance, it may be very costly in a large, complex society to maintain an institution that excludes nontaxpayers from driving on public roads. On the other hand, in a small village in which neighbors often help repair one another's homes, an individual who refuses to help can be ostracized relatively easily: When his home needs repairing, his neighbors can refuse to help him.

### *The Fork in the Road*

The rejection of positive assortment as an evolutionary explanation of collective action has been based mainly on casually formulated and informally defended assumptions about the abilities of cooperators to identify one another and to ostracize free riders. Reciprocal altruism, on the other hand, has been rejected somewhat more rigorously. This rejection can in fact be traced back to a specific influential publication, "The Evolution of Reciprocity in Sizable Groups" (Boyd & Richerson, 1988), which concluded that reciprocity could allow for the evolution of only very small



collective actions and that it is an inadequate explanation for the large-scale collective actions in which humans interact. Boyd and Richerson's (1988) paper was an important "fork in the road" that led many researchers to veer away from individual-level adaptationist accounts of human cooperation. Since Boyd and Richerson, the tendency has been for evolutionary modelers to simply note that reciprocity does not apply directly to collective action contexts and to proceed to explore alternative explanations instead—alternatives that do not invoke Trivers' theory (Brandt, Hauert, & Sigmund, 2003; Fehr, 2004; Henrich, 2004; Panchanathan & Boyd, 2004), and which often depend on group selection (Boyd et al., 2003; Fehr & Fischbacher, 2003; Gintis, 2000). Given the influential role that Boyd and Richerson have had in rejecting individual-level adaptationist accounts of collective action, it is worth revisiting this publication and examining its conclusions in some detail (see also Trivers, 2006).

First, let us assume that there is a population with two strategies, *R* (reciprocator) and *F* (free rider). Boyd and Richerson (1988) defined reciprocity, *R*, in the collective action as a strategy which cooperates as long as all the other  $n - 1$  group members cooperate, but stops cooperation immediately if one or more group members free ride. (The tolerance of *R*, i.e., the number of comembers who must cooperate in order for *R* to cooperate, can vary in Boyd and Richerson's model, but for now we will assume that this value is  $n - 1$ ). This means that *R* continues cooperating only when all its comembers are also *Rs*. In a mixed group with *R* and *F*, *R* cooperates in the first round and then stops cooperating from the second round. *Rs* are exploited in such mixed groups but only in the first round. When *R* is in a group consisting only of other *Rs*, it can enjoy the benefit of sustained mutual cooperation. Thus, whether *R* can avoid extinction and increase its share (i.e., its relative proportion) in a population depends on how likely *R* is to find itself in a group consisting of only *Rs*.

Second, let us assume that multiple groups of  $n$  individuals are randomly formed and members play a game of public goods (i.e., they interact in a collective action where  $n > m > 1$ , as described previously) in each group for a certain number of rounds. This process of random matching is a standard assumption in most evolutionary game theoretical models of cooperation. As there are only two strategies in this population, group composition simply follows a binomial process. In principle, if the average payoff to *R* exceeds that of *F*, *R* can evolve and increase its share in the population. According to Boyd and Richerson (1988), "[R]eciprocity is likely to evolve only when reciprocating groups are quite small" (p. 352) simply because it is very challenging for *Rs* to form groups of only *Rs*, even when *R* is common in the population. For example, consider a population with 80% of *Rs* and 20% of *Fs*, and in which groups of 3 members are randomly formed. The probability of a group consisting of 3 *Rs* is  $0.8^3 (= 0.512)$ . If group size increases to 20, however, this probability quickly decreases to 0.012. As group size gets larger, this probability decreases exponentially, and it becomes increasingly difficult for *R* to reap the benefits of mutual cooperation. Thus, Boyd and Richerson conclude, "[T]he conditions that allow the evolution of reciprocal altruism become extremely restrictive as group size increases" (p. 337).

If *R* gets more tolerant by continuing to cooperate even when more of its comembers free ride, for instance, if it cooperates as long as  $n - 2$  of its comembers cooperate (instead of  $n - 1$ , as with the previous example), then *R* will have more opportunities to benefit from mutual cooperation. However, the more tolerant *R* becomes, the more *F* can exploit the cooperation enabled by *R*'s tolerance, and Boyd and Richerson (1988) show that increasing *R*'s tolerance does not allow *R* to overcome *F*. And while positive assortment can increase the probability that *R* will find itself in groups that include other *Rs* and that exclude *F*, Boyd and Richerson show that even with the help of very strong positive assortment, *R*'s evolution still quickly becomes difficult as group size increases.

Boyd and Richerson (1988, p. 352) suggest that collective actions involving 3 individuals could evolve relatively easily even with low positive assortment (i.e., the probability that a reciprocator's comember will also be a reciprocator is modified and increased by a parameter  $r = 1/32$ , which is comparable to Hamilton's kin coefficient), and with few (i.e., about 10) repeated interactions. However, in collective actions involving 16 individuals, even with high positive assortment

( $r = 1/4$ ), an unrealistically high number of interactions (about 10 million) would be required in order for reciprocity to evolve. Critically, subsequent publications stated that the Boyd and Richerson model shows that reciprocity could evolve in collective actions of “about 10” (Henrich, 2004, p. 10) or “a handful of” (Boyd et al., 2003, p. 3531) participants. These authors note that human collective actions are often much larger than this and, therefore, conclude that such cooperation cannot be explained in terms of reciprocal altruism. There are, however, a number of problems with this rejection of reciprocity as an explanation for large-scale cooperation, to which we now turn.

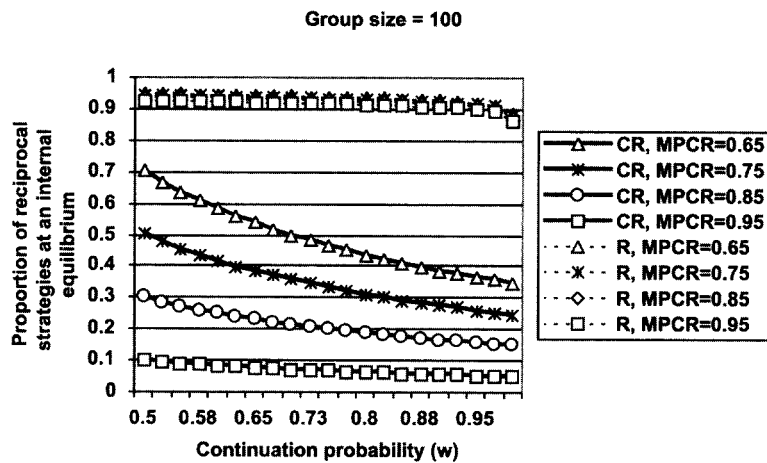
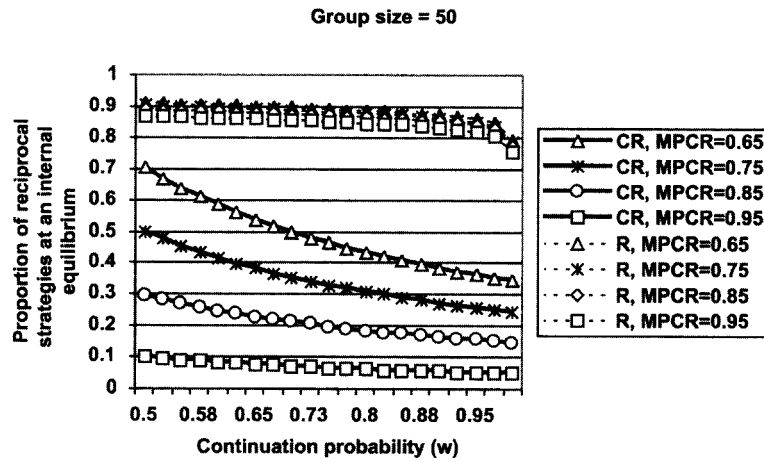
### *Proximate Causation*

First, even if psychological mechanisms for reciprocity did evolve in small groups, they could still produce reciprocity in larger ones. Hunter-gatherer bands typically consist of about 7–8 male and female full-time foragers (Kelly, 1995, p. 213), and the average sex-segregated foraging group may involve only 3–5 members. If reciprocators in these groups followed a scalable decision rule similar to “contribute to the observed (or expected) extent of the average co-participant,” then they might have been motivated to reciprocate in larger groups as well (the larger the group, the larger the denominator used to calculate the average contribution). Even if selection against such a strategy would occur in modern environments characterized by large collective actions, it could be that when people in industrialized societies participate in large-scale cooperation (e.g., vote in elections, fight in wars), they are motivated by proximate adaptations that evolved by promoting reciprocity in much smaller groups.

### *Replacing Binary Reciprocity with Continuous Reciprocity*

The previous argument about proximate causation is not a direct criticism of Boyd and Richerson’s (1988) model, but simply raises the question of whether reciprocity should be rejected as an explanation of collective action, based on the results of a single theoretical model. In this section, we provide a more fundamental criticism of Boyd and Richerson’s rejection of reciprocity by pointing out and reexamining one typical error that theoreticians tend to commit. Boyd and Richerson’s model explored only a binary reciprocal strategy that contributes either all (1.0) or nothing (0.0) to the public good. However, a more realistic model would allow for *continuous* reciprocity, that is, for contributions that can vary continuously between 0.0 and 1.0. Binarization of individual behavior is a simplification of a more complex reality that may sometimes be necessary for making a model tractable. Further, the difference between a binary and continuous reciprocal strategy may seem so insignificant that one may be tempted to conclude that Boyd and Richerson’s simplification is acceptable. However, in the following section, we show that if Boyd and Richerson had looked at continuous rather than binary reciprocity, then they may have reached a quite different conclusion about the adaptiveness of reciprocity in sizable groups (and the subsequent history of research on human cooperation might have been quite different, too).

Consider a new strategy, *CR*, that contributes in round  $t$  as a function of the average contribution made by the  $n - 1$  other group members in round  $t - 1$  (i.e., in the previous round). Takezawa and Price (2006) consider a single variant of *CR* that simply matches the average comember contribution from the previous round, and their model is in all other respects identical to the model presented in Boyd and Richerson (1988). Figure 17.1 shows the minimum proportion of reciprocal strategies (*R* and *CR*) that is necessary for reciprocators to spread at the expense of *F*. *R*’s share in a population needs to be very high—above 80%—when group size is 50. This value must be high, regardless of the value of several other parameters in the model, such as the efficiency of cooperation (marginal per capita return, or *MPCR*—a value that is higher in groups that produce more resources for a given level of effort), and regardless of the length of the interaction ( $w$ ; Figure 17.1a). In contrast, *CR* can increase its share against *F* even when its initial proportion in the population is relatively low, and especially when cooperation is very efficient. For example when *MPCR* = 0.95, *CR* can evolve even when its initial share is only 5% of a population. Figure 17.1b indicates a case



**Figure 17.1a–b** The minimum proportion of reciprocal strategies (*R* or *CR*) necessary to invade a population of free riders (*F*).

where group size is 100, and continues to show a similar pattern: *CR* can evolve even when it is rare, as long as cooperation is efficient.

Next we will consider a third cooperative strategy, the unconditional cooperator, *C*. This strategy contributes fully in every round, regardless of how its comembers behave. The bad news for *C* is that in a population consisting of only *C* and *F*, even a single *F* can invade and take over, because there is nothing that *C* can do to defend itself from being exploited to extinction. The bad news for *R*, however, is that when *F* is absent in a population, *C* will not be exploited and so can receive the same payoff as *R*. If *C* is rare and it invades a population of *CR*, stochastic processes such as mutation may increase the share of *C*. As the proportion of *C* increases, however, *F* may be able to invade the population by exploiting *C*, which may result in the extinction of both *C* and *CR*. It is therefore important to check the robustness of *CR*, assuming that there are three strategies in a population—*F*, *C*, and *CR*.

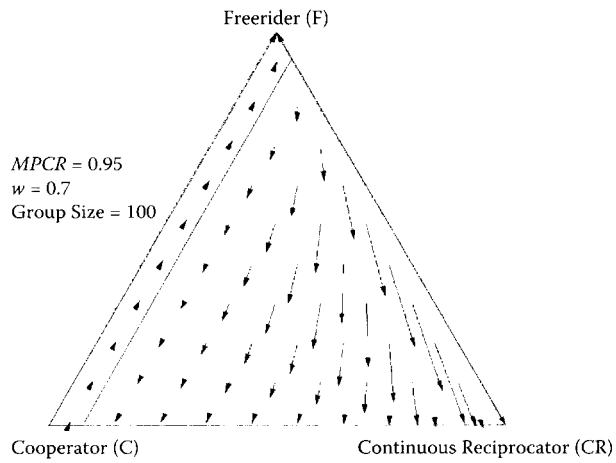
In numerical computer simulations, we varied the proportions of the three strategies *F*, *C*, and *CR*, and found that the system evolved toward a mixture of *C* and *CR* in most combinations of proportions of the three strategies. Figure 17.2a is called a ternary plot. Any one point within this equilateral triangle represents the relative proportions of the three strategies in a population, which sum to 100%. Each base (or side) of the triangle represents a minimum proportion of 0% of a given strategy, while the point opposite represents a maximum proportion of 100% of that strategy. For instance, if a point is near the bottom-right corner, it means that a population is dominated by *CR*. Any proportion between these limits, for all three strategies, can thus be simultaneously represented by a single point within the triangle. For example if the point is somewhere along the perimeter between the top corner and bottom-right corner, it means that only two strategies, *CR* and *F*, exist in a population (and *C* is absent). The arrow at each point indicates the *direction* in which the system is evolving, for instance an arrow pointing toward the bottom-right corner indicates that the system is evolving toward domination by *CR*. The length of the arrow indicates the *strength* of evolution in that direction. Finally, the line that bisects the triangle divides it into the area in which the system evolves toward a *C-CR* mixture and that which evolves toward a pure equilibrium of *F*. In Figure 17.2a, the *C-CR* area is much larger than the *F* area. Roughly speaking, the cooperative strategies *C* and *CR* can evolve as long as the proportion of *CR* is larger than 5–6% under the parameter combination  $MPCR = 0.95$ ,  $w = 0.7$ , group size = 100. Very similar results are found even when group size is increased to 500 (Figure 17.2b). Figure 17.2c shows a ternary plot where *CR* is replaced by Boyd and Richerson's (1988) binary reciprocal strategy *R*, and shows that the area leading to the evolution of cooperation is much smaller than when reciprocity is continuous.

Why did replacing binary reciprocity with continuous reciprocity generate such a dramatic difference? Consider a group of five individuals. As we discussed in the previous section, *R* can gain the benefits of mutual cooperation only when all four of its comembers are also *R*. Just one *F* in the group induces the collapse of cooperation. On the other hand, when four *CR* and one *F* exist in a group, *CR* still can enjoy some degree of mutual cooperation (Table 17.2a). When the number of *F* increases, for instance to three, *CR* stops cooperating very quickly and can avoid being exploited by *F* (Table 17.2b). The power of *CR* is in its capacity to quickly reduce cooperation when there are too many *F*s among its comembers, while maintaining a moderately high level of cooperation when there are fewer *F*s.

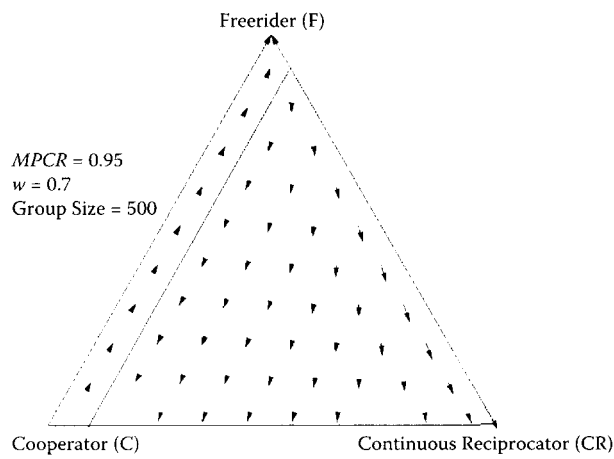
### *Some Notes on Interpretation of the Model Results*

Although quite different conclusions can be drawn from a model of continuous as opposed to binary reciprocity, one must be careful in interpreting these results. First, the area favoring the evolution of cooperative strategies in the ternary plot gets smaller quickly as the efficiency of cooperation, *MPCR*, goes down (see also Figure 17.1). The *MPCR* value of 0.95, which is used in Figure 17.2, is very high compared to the values used in many public goods game experiments, which are usually around 0.40 (though note that we used the *same* *MPCR* value when comparing our *CR* model with Boyd and Richerson's *R* model). Unless evidence is provided that collective actions in human ancestral environments were very efficient, it should not be concluded that our model is sufficient for explaining the evolution of large-scale cooperation. Second, even when cooperation is very efficient, at least 5–6% of the population needs to mutate into *CR* at the same time in order for *CR* to overcome domination by *F*. If stochastic forces are sufficient to make such a multiple mutation realistic, the system starts to drift and may oscillate between an equilibrium of all *F* or a mixture of *C-CR*, in which case one must examine in which equilibrium, *F* or *C-CR*, the system spends more time.

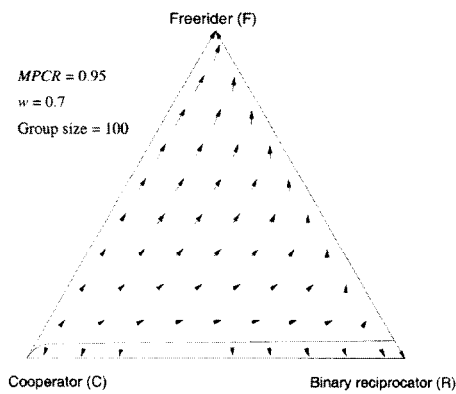
Despite these issues, it is noteworthy that this model challenges Boyd and Richerson's (1988) argument that reciprocity cannot explain the evolution of cooperation, and does not rely on any exotic assumptions such as group extinctions (Boyd et al., 2003; Gintis, 2000) or cultural group selection (Boyd & Richerson, 1985; Henrich, 2004).



(a)



(b)



(c)

**Figure 17.2a–c** Dynamics of unconditional cooperators (*C*), free riders (*F*) and reciprocal (*CR*) strategies, for continuous reciprocity under two sets of parameter values (a and b), and when reciprocity is binary instead of continuous as in Boyd and Richerson’s (1988) original model (c).

**Table 17.2a–b Transition of the Amount of Contributions by Continuous Reciprocators across Six Rounds When There Are: (a) Four Continuous Reciprocators and One Free Rider and (b) Two Continuous Reciprocators and Three Free Riders**

(a)

	Rounds					
	1	2	3	4	5	6
CR	1	0.75	0.56	0.42	0.32	0.24
CR	1	0.75	0.56	0.42	0.32	0.24
CR	1	0.75	0.56	0.42	0.32	0.24
CR	1	0.75	0.56	0.42	0.32	0.24
F	0	0	0	0	0	0

(b)

	Rounds					
	1	2	3	4	5	6
CR	1	0.25	0.06	0.02	<0.01	<0.01
CR	1	0.25	0.06	0.02	<0.01	<0.01
F	0	0	0	0	0	0
F	0	0	0	0	0	0
F	0	0	0	0	0	0

### *Rethinking Positive Assortment*

As previously demonstrated, a single rather simple adjustment to Boyd and Richerson's (1988) model—changing the reciprocity strategy from binary to continuous—produces a dramatically different result. Compared to the binary strategy, continuous reciprocity is much better at replicating when it is rare in a population. Further, reciprocity's enhanced adaptiveness is achieved even without assuming any higher a level of positive assortment among reciprocators than assumed in the original model. When reciprocators are allowed to positively assort, *synergy* between the two behaviors may further enhance reciprocity's ability to evolve in groups. The effect found in our model, therefore, may be on the conservative side.

The fact that positive assortment can interact synergistically with other cooperative strategies points to another problem that has plagued models of the evolution of cooperation—*population structure* (or lack thereof). While reciprocity may not survive in populations that intermix randomly, as in Boyd and Richerson (1988), if individuals can *choose* their interaction partners and preferentially cooperate with them, then reciprocity can survive even in large populations. Models that assume *random mixing* of interactions are misleading because they artificially exaggerate the effective population size. Even in a big city like New York, one tends to interact with a small subset of the same people every day—family, colleagues, bosses, friends, and so forth. In real life, choosing interaction partners means that the likelihood of meeting again remains high over time, because individual A will *seek out* individual B and vice versa. Unless one forces them to rely on bumping

into each other at some random location in New York each time, A can simply go and find B (e.g., go and knock on their door).

Another potentially misleading assumption is that the population is mixed up so that individuals must find each other at all. Most human groups have a number of *structural features* that ensure interactions are *constrained* to certain dyads and groups and not others (over and above any effects of preference). These include: spatial structure (one is more likely to interact with neighbors or people one passes regularly), hierarchies (one is more likely to interact with certain ranks within the hierarchy and not others), peer groups (one is more likely to interact with members who have similar characteristics), and activities (one is more likely to interact with groups formed for specific reasons, e.g., hunting parties, warrior groups, etc., which do not constitute the whole population).

The dynamics of positive assortment are absolutely crucial to constructing realistic models of human social behavior *and* to understanding the paths not taken in the cooperation literature. Low levels of positive assortment can drastically increase cooperation, as shown in many spatial prisoner's dilemma games (Axelrod, 1997; Nowak, May, & Sigmund, 1995; Hammond & Axelrod, 2006; Nowak, 2006), as well as accelerate the spread of norms within structured networks, rendering group selection models "neither justified nor necessary. Such behaviors can emerge from individual-based models, simply involving rewards to individuals who belong to groups" (Ehrlich & Levin, 2005, p. 946; see also Durrett & Levin, 2005; Nakamaru & Levin, 2004).

To sum up, previous models have assumed that people reciprocate according to an all or nothing binary rule, and that people either are not allowed to choose who they interact with at all, or they have to rely on chance. Both are clearly inaccurate descriptions of human social behavior. Allowing people to reciprocate continuously, and allowing them to choose who they interact with, is far more realistic and, crucially, leads to radically different predictions about whether or not cooperation in groups could evolve via individual-level adaptation.

### WHERE DOES THE EVIDENCE POINT? SUPPORT FOR THE INDIVIDUAL-LEVEL ADAPTATIONIST VIEW

We argued previously that human collective action could in theory be the product of individual-level adaptation, if evolution favored reciprocity in collective actions and/or positive assortment among cooperators. As well as exploring such theoretical issues, another avenue of inquiry is to assess whether existing models accurately describe how people actually behave in real life. Does the empirical evidence point toward group selection, or do members of cooperative groups behave as predicted by the individual-level theories just discussed? We address these questions in this section, for each of several topics in turn.

#### *Conditional Cooperation in Public Good Games*

A standard finding of experimental collective actions (public good games), is that the majority of subjects are "conditional" cooperators, that is, they cooperate more when they perceive that coplayers are more willing to cooperate (Fischbacher, Gächter, & Fehr, 2001; Kurzban & Houser, 2005; Ledyard, 1995), and less when they believe that coplayers are free riding (Fehr & Gächter, 2000; Kurzban, et al., 2001). "Conditional cooperation" in these games, then, is essentially a term for reciprocity between a player and multiple coplayers. Anthropologists have long considered reciprocity to be a cross-cultural universal (Brown, 1991; Gouldner, 1960), and conditional cooperation in collective actions may be similarly universal. Existing cross-cultural data on this point are scant but supportive: In a public good game conducted among the Mapuche of southern Chile, players' contribution amounts were best predicted by their beliefs about how much coplayers would contribute (Henrich & Smith, 2004). Research on the neurological basis of reciprocity may provide clues about the proximate motivations behind conditional cooperation in public good games. In one fMRI study, reciprocity was associated with consistent activation in brain areas that have been linked

with reward processing (Rilling et al., 2002). This suggests that reciprocity has deeply engrained biological roots.

Consistent with our previous model that represents reciprocity as a continuous rather than binary strategy, cooperators in public good games apparently strive to match the average coplayer contribution. Experimental studies (Fischbacher et al., 2001; Kurzban & Houser, 2005) suggest that one half to two thirds of public good game subjects engage in a reciprocal strategy whereby their contributions correlate positively with the average observed or expected cosubject contribution. Another study (Croson, 1999) found that players' contributions were better predicted by their expectations about the mean coplayer contribution than by their expectations about the minimum or maximum coplayer contribution. Further, players' contributions were predicted not just by the *mean expected* coplayer contribution but by the *actual mean* coplayer contribution; in other words, players' expectations tended to be accurate. Another public good study (Keser & van Winden, 2000) found that subjects tended to increase contributions after discovering that they had been contributing less than the average coplayer, decrease contributions after discovering that they had been contributing above the average and maintain contribution levels after discovering that they had been matching the average.

### *Signaling Cooperativeness Initially and the Decay of Cooperativeness Over Repeated Rounds*

Evidence just reviewed suggests that public good game subjects tend to match expected coplayer contributions. But what if subjects are interacting for the first time, and lack information on which to base expectations about coplayer contributions? In that case, conditional cooperators would need to be biased toward initial cooperativeness, in order to signal their cooperativeness to each other and, perhaps, prime the interaction with a high initial average contribution. Thus, when interactants have no information about each other's cooperativeness, they should be biased toward cooperativeness at the start of the interaction (analogous to the optimal strategy in iterated prisoner's dilemma, Axelrod, 1984). All else equal, they should then continue to cooperate to the extent that they perceive cointeractants to be reciprocating. And indeed, a standard finding of anonymous public good games, in which players have no information about coplayer cooperativeness at the start of the game, is that cooperative behavior follows the pattern described previously. Contributing is highest at the outset of the game and, then, decays gradually over time, as players receive information that some coplayers are free riding (Ledyard, 1995; Masclet, Noussair, Tucker, & Villeval, 2003). This gradual decay apparently occurs because higher contributors in initial rounds down-ratchet their contributions as the game progresses, in order to match the average expected coplayer contribution, which dwindles constantly due to a persistent minority of free riders (Fischbacher et al., 2001; Kurzban et al., 2001).

### *Mutual Monitoring*

If people cooperate to the extent to which they expect coparticipants to reciprocate, then they must continually monitor coparticipants in order to assess their level of reciprocation. Evidence indicates that cross-culturally, successful collective actions involve mutual monitoring efforts by participants (Erasmus, 1977; Ostrom, 1990). Interactants appear to monitor one another not just constantly but accurately. In a Shuar hunter-horticulturalist village, *perceptions* of covillagers' engagement in provillage altruism correlated positively with covillagers' *actual* engagement in specific altruistic activities (Price, 2003). In a sugarcane cultivating workgroup in this same society, perceptions of coworker cooperativeness (attendance record and physical effort in work sessions) correlated positively with more objective measures of this cooperativeness (Price, 2006). Moreover, participants accurately distinguished "intentional" low contributors (those who could have contributed highly but chose not to) from "unintentional" low contributors (those physically *unable* to contribute highly). Experimental studies provide additional evidence that people are cognitively adapted for



mutual monitoring. Cross-culturally, subjects detect violators of social contracts quickly and easily: Both Western and Amazonian Shiwiar subjects are better at solving a logic problem when the problem is presented in terms of detecting a cheater in a social exchange (Cosmides & Tooby, 1992; Sugiyama, Tooby, & Cosmides, 2002). A brain-injured patient showed selective impairment in this cheater-detection ability (Stone, Cosmides, Tooby, Kroll, & Knight, 2002), suggesting that the ability is enabled by a specialized cognitive mechanism. As Martin Nowak and Karl Sigmund (2005) recently concluded, "The moralistic assessment of the other members in the population, even if they are observed only at a distance, provides a powerful tool for channelling support towards those who collaborate, and an incentive to join group efforts" (p. 1296).

### *Positive Assortment Via Partner Choice*

Experimental evidence suggests that both cooperators and free riders choose partners whom they expect will cooperate more. When subjects are allowed partner choice, cooperators attempt to interact with other cooperators and to avoid free riders, while free riders attempt to avoid being avoided. In a study that allowed subjects to join the public good game group of their choosing, Ehrhart and Keser (1999) observed a continual flight of the more cooperative subjects away from the less cooperative ones, as the more cooperative subjects attempted to form groups with each other, and the less cooperative subjects attempted to join groups of cooperators in order to free ride. In another public good game study, players were informed of coplayers' contribution histories and were allowed to form cooperative groups based on mutual partner preference. The highest contributors chose to cooperate with each other and proceeded to interact more productively than less cooperative coplayers did (Page, Putterman, & Unel, 2005). Similarly, when Sheldon, Sheldon and Osbaldiston (2000) allowed subjects to form groups with the comembers of their choosing, they found that relatively prosocial individuals tended to cluster together and cooperate relatively productively. Preferential interaction with expected cooperators has also been elicited experimentally in the form of indirect reciprocity: Experimental subjects will cooperate more with people whom they have observed cooperating with others (Barclay, 2004; Milinski, Semmann, & Krambeck, 2002; Nowak & Sigmund, 2005; Wedekind & Milinski, 2000) and whom they may therefore expect would cooperate with them. Further, the finding that cooperation flourishes in environments characterized by high levels of "trust," that is, environments in which people expect others to cooperate, has been reported consistently in the literature of organizational behavior and other fields (Dirks & Ferrin, 2001; Jones & George, 1998; Kramer, 1999; Ostrom, 2003).

### *Reputation*

Humans universally acquire reputations based on their cooperativeness (Alexander, 1987; Panchanathan & Boyd, 2003; Trivers, 1971). Numerous studies suggest a correlation between cooperativeness and goodness of reputation in small scale societies, with reputations being based on, for example, how much people provision resources such as food (Gurven, Allen-Arave, Hill, & Hurtado, 2000; Hawkes, 1993), military service (Chagnon, 1988; Meggitt, 1977), and labor in group work projects (Price, 2003, 2006). An altruism-reputation correlation was also reported in a study of a large western business firm, in which employees' social status was positively related to their perceived generosity (Flynn, 2003). Further, experimental results suggest that people cooperate more when they can acquire a reputation (Barclay, 2004; Milinski et al., 2002). The importance of reputation effects for sustaining cooperation has also been noted in fields outside of experimental economics, such as transaction cost economics (Demsetz, 1993; Williamson, 1996). The selection of cooperative partners may be enabled by psychological mechanisms for remembering potential interactants based on their reputations. When subjects were shown photographs of people and given various kinds of information about them and then asked several days later which of them they remembered, subjects displayed enhanced memory for those who had been portrayed as cheaters in cooperative interactions (Mealey, Daood, Krage, 1996; Oda, 1997). Further, compared to neutral faces, faces

of people who have intentionally cooperated or defected in a prisoner's dilemma game are remembered better and elicit more activity within a neural system associated with social cognition (Singer, Kiebel, Winston, Dolan, & Frith, 2004).

The previous examples assume that assessment of others' cooperativeness would be based on information about others' actual past engagement in cooperative behavior. However, cooperativeness also seems to be assessed via information gathered during verbal communication and face-to-face interaction. Subjects cooperate more when they have been given opportunities for such interaction (for a review, see Ostrom, 2003), perhaps because they can discuss their intentions (Ostrom, Walker, & Gardner, 1992), and make inferences about coplayer dispositions (Frank, Gilovich, & Regan, 1993), facial expressions and features (Brown & Moore, 2002; W. Brown, Palameta, Moore, 2003; Eckel & Wilson, 2003; Yamagishi, Tanida, Mashima, Shimoma, & Kanazawa, 2003), or other verbal and nonverbal indicators of cooperative intent (Bochet, Page, & Putterman, in press; Brown et al., 2003).

## CONCLUSIONS

The mainstream literature on the evolution of cooperation has recently tended toward a group-selectionist view. However, as we have demonstrated in this chapter, progress in understanding human cooperation requires not the abandonment of individual-level adaptationist theories, but rather the extension and synthesis of these theories. The group selectionist paradigm faces significant theoretical and empirical challenges (Burnham & Johnson, 2005; Hagen & Hammerstein, 2006; Haley & Fessler, 2005; Price et al., 2002), while individual-level theories appear to be superior at predicting how collective action participants actually behave. There is good evidence that humans (and perhaps some animals, Stevens & Hauser, 2004) behave as reciprocators and engage in positive assortment. Both behaviors are predicted directly by the individual-level theories and the model described above but *not* by theories of biological and cultural group selection.

The important challenge facing students of human cooperation in the future is to identify the key factors that influence variation in human cooperative behavior and to design experiments that expose the evolved psychological mechanisms that regulate this behavior. Another challenge remains for the group-selectionists: to provide *any* empirical evidence for their favored view of cooperation. A considerable and growing body of empirical work has begun to reveal genetically selfish motives—serving the interests of individuals rather than groups—underlying so-called “strong reciprocity” (Bateson et al., 2006; Burnham & Hare, in press; Burnham, 2003; Haley & Fessler, 2005; Kurzban, DeScioli, & O'Brien, 2007). It will be interesting to see if the group selectionists can maintain the direction of their march (to which they have already committed themselves theoretically) against the rising tide of empirical evidence for individual-level adaptations for cooperation.

It is ironic that one of the most basic and everyday features of human nature—cooperation—is so poorly understood by behavioral scientists, despite a plethora of great minds working on and discussing the problem in the pages of the world's leading science journals. What is more, it is urgent that the problem is resolved. If we cannot agree on even the basic origins and motives of human cooperation, how can we hope to promote it in our rapidly globalizing world of environmental devastation, genocide, terrorism, poverty, and war? The mainstream literature has, we believe, been led astray largely because early models presented an erroneous “fork in the road,” which caused the majority of theorists to discount the power of reciprocity to explain human cooperation in groups and to rush headlong into group selection. As Robert Frost might have advised us, where those two roads diverge, we would now be wise to explore the one less travelled.

## REFERENCES

- Alexander, R. D. (1987). *The biology of moral systems*. Aldine, NY: Hawthorne.
- Axelrod, R. (1984). *The evolution of cooperation*. London: Penguin.
- Axelrod, R. (1997). *The complexity of cooperation: Agent-based models of competition and collaboration*. Princeton, NJ: Princeton University Press.
- Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, 211, 1390–1396.
- Barclay, P. (2004). Trustworthiness and competitive altruism can also solve the “tragedy of the commons.” *Evolution and Human Behavior*, 25, 209–220.
- Barkow, J. H., Cosmides, L., & Tooby, J. (Eds.). (1992). *The adapted mind: Evolutionary psychology and the generation of culture*. Oxford, U.K.: Oxford University Press.
- Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters*, 2, 412–414.
- Bochet, O., Page, T., & Putterman, L. (2006). Communication and punishment in voluntary contribution experiments. *Journal of Economic Behavior and Organization*, 60(1), 11–26.
- Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences*, 100, 3531–3535.
- Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process*. Chicago: University of Chicago Press.
- Boyd, R., & Richerson, P. J. (1988). The evolution of reciprocity in sizable groups. *Journal of Theoretical Biology*, 132, 337–356.
- Brandt, H., Hauert, C., & Sigmund, K. (2003). Cooperation, punishment and reputation in spatial games. *Proceedings of the Royal Society, London B*, 270, 1099–1104.
- Brown, D. E. (1991). *Human universals*. New York: McGraw-Hill.
- Brown, W. M., & Moore, C. (2002). Smile asymmetries and reputation as reliable indicators of likelihood to cooperate: An evolutionary analysis. *Advances in Psychology Research*, 11, 59–78.
- Brown, W. M., Palameta, B., & Moore, C. (2003). Are there nonverbal cues to commitment? An exploratory study using the zero-acquaintance video presentation paradigm. *Evolutionary Psychology*, 1, 42–69.
- Burnham, T. C., & Hare, B. (in press). Engineering cooperation: Does involuntary neural activation increase public goods contributions? *Human Nature*.
- Burnham, T. C., & Johnson, D. D. P. (2005). The biological and evolutionary logic of human cooperation. *Analyse & Kritik*, 27(1), 113–135.
- Burnham, T. C. (2003). Engineering altruism: A theoretical and experimental investigation of anonymity and gift giving. *Journal of Economic Behavior and Organization*, 50(1), 133–144.
- Chagnon, N. A. (1988). Life histories, blood revenge, and warfare in a tribal population. *Science*, 239, 985–992.
- Chagnon, N. A., & Irons, W. (Eds.). (1979). *Evolutionary biology & human social behaviour: An anthropological perspective*. North Scituate, MA: Duxburg Press.
- Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163–228). New York: Oxford University Press.
- Croson, R. T. A. (1999). *Contributions to public goods: Altruism or reciprocity?* Philadelphia: Wharton School of Business, University of Pennsylvania.
- Darwin, C. (1871). *The descent of man*. New York: Penguin Classics.
- Dawkins, R. (1976). *The selfish gene*. Oxford, U.K.: Oxford University Press.
- Demsetz, H. (1993). The theory of the firm revisited. In O. E. Williamson & S. G. Winter (Eds.), *The nature of the firm: Origins, evolution and development* (pp. 159–178). New York: Oxford University Press.
- Dirks, K. T., & Ferrin, D. L. (2001). The role of trust in organizational settings. *Organization Science*, 12, 450–467.
- Dugatkin, L. A. (1997). *Cooperation in animals*. Oxford, U.K.: Oxford University Press.
- Durrett, R., & Levin, S. A. (2005). Can stable social groups be maintained by homophilous imitation alone? *Journal of Economic Behavior and Organization*, 57(3), 267–286.
- Eckel, C. C., & Wilson, R. K. (2003). The human face of game theory: Trust and reciprocity in sequential games. In E. Ostrom & J. Walker (Eds.), *Trust and reciprocity: Interdisciplinary lessons from experimental research* (pp. 245–274). New York: Sage.

- Ehrhart, K., & Keser, C. (1999). *Mobility and cooperation: On the run* (pp. 99s–24). Montreal, Quebec, Canada: CIRANO.
- Ehrlich, P., & Levin, S. (2005). The evolution of norms. *PLoS Biology*, 3(6), e194.
- Erasmus, C. J. (1977). *In search of the common good*. New York: The Free Press.
- Fehr, E. (2004). Don't lose your reputation. *Nature*, 432, 449–450.
- Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, 425, 785–791.
- Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90, 980–994.
- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415, 137–140.
- Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, 71, 397–404.
- Flynn, F. J. (2003). How much should I give and how often? The effects of generosity and frequency of favor exchange on social status and productivity. *Academy of Management Journal*, 46, 539–553.
- Frank, R. H., Gilovich, T., & Regan, D. T. (1993). The evolution of one-shot cooperation: An experiment. *Ethology and Sociobiology*, 14, 247–256.
- Gadagkar, R. (2001). *Survival strategies: Cooperation and conflict in animal societies*. Cambridge, MA: Harvard University Press.
- Gintis, H. (2000). Strong reciprocity and human sociality. *Journal of Theoretical Biology*, 206, 169–179.
- Gintis, H. (2003). Solving the puzzle of prosociality. *Rationality and Society*, 15, 155–187.
- Gintis, H., Smith, E., & Bowles, S. (2001). Costly signalling and cooperation. *Journal of Theoretical Biology*, 213, 103–119.
- Gouldner, A. W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review*, 25, 161–178.
- Curven, M., Allen-Arave, W., Hill, K., & Hurtado, A. M. (2000). 'It's a wonderful life': Signaling generosity among the Ache of Paraguay. *Evolution and Human Behavior*, 21, 263–282.
- Hagen, E. H., & Hammerstein, P. (2006). Game theory and human evolution: A critique of some recent interpretations of experimental games. *Theoretical Population Biology*, 69, 339–348.
- Haig, D. (1996). Gestational drive and the green-bearded placenta. *Proceedings of the National Academy of Sciences*, 93, 6547–6551.
- Haley, K., & Fessler, D. (2005). Nobody's watching? Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior*, 26, 245–256.
- Hamilton, W. D. (1964). The genetical evolution of social behaviour, I & II. *Journal of Theoretical Biology*, 7, 1–52.
- Hammond, R. A., & Axelrod, R. (2006). The evolution of ethnocentrism. *Journal of Conflict Resolution*, 50(6), 926–936.
- Hawkes, K. (1993). Why hunter-gatherers work—An ancient version of the problem of public goods. *Current Anthropology*, 34, 341–361.
- Henrich, J. (2004). Cultural group selection, coevolutionary processes and large-scale cooperation. *Journal of Economic Behavior & Organization*, 53, 3–35.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., & Gintis, H. (Eds.). (2004). *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*. Oxford, U.K.: Oxford University Press.
- Henrich, J., & Smith, N. (2004). Comparative experimental evidence from machiguenga, mapuche, huinca & American populations. In J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr, & H. Gintis (Eds.), *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies* (pp. 125–167). Oxford, U.K.: Oxford University Press.
- Johnson, D. D. P., Stopka, P., & Knights, S. (2003). The puzzle of human cooperation. *Nature*, 421, 911–912.
- Jones, G. R., & George, J. M. (1998). The experience and evolution of trust: Implications for cooperation and teamwork. *Academy of Management Review*, 23, 531–546.
- Keller, L. (2002). Marsupial sperm pairing: A case of 'sticky' green beards? *Trends in Ecology and Evolution*, 17, 113.
- Keller, L., & Ross, K. G. (1998). Selfish genes: A green beard in the red fire ant. *Nature*, 394, 573–575.
- Kelly, R. L. (1995). *The foraging spectrum: Diversity in hunter-gatherer lifeways*. Washington, DC: Smithsonian Institution Press.

- Keser, C., & van Winden, F. (2000). Conditional cooperation and voluntary contributions to public goods. *Scandinavian Journal of Economics*, *102*, 23–29.
- Kramer, R. M. (1999). Trust and distrust in organizations: Emerging perspectives, enduring questions. *Annual Review of Psychology*, *50*, 569–598.
- Kurzban, R., DeScioli, P., & O'Brien, E. (2007). Audience effects on moralistic punishment. *Evolution and Human Behavior*, *28*, 75–84.
- Kurzban, R., & Houser, D. (2005). Experiments investigating cooperative types in humans: A complement to evolutionary theory and simulations. *PNAS*, *102*(5), 1803–1807.
- Kurzban, R., McCabe, K., Smith, V. L., & Wilson, B. J. (2001). Incremental commitment and reciprocity in a real time public goods game. *Personality and Social Psychology Bulletin*, *27*, 1662–1673.
- Ledyard, J. O. (1995). Public goods: A survey of experimental research. In J. H. Kagel & A. E. Roth (Eds.), *The handbook of experimental economics* (pp. 111–194). Princeton, NJ: Princeton University Press.
- Masclet, D., Noussair, C., Tucker, S., & Villeval, M. (2003). Monetary and nonmonetary punishment in the voluntary contributions mechanism. *American Economic Review*, *93*, 366–380.
- Mealey, L., Daood, C., & Krage, M. (1996). Enhanced memory for faces of cheaters. *Ethology and Sociobiology*, *17*, 119–128.
- Meggitt, M. (1977). *Blood is their argument*. Palo Alto, CA: Mayfield.
- Milinski, M., Semmann, D., & Krambeck, H. (2002). Reputation helps solve the “tragedy of the commons.” *Nature*, *415*, 424–426.
- Nakamaru, M., & Levin, S. A. (2004). Spread of two linked social norms on complex interaction networks. *Journal of Theoretical Biology*, *230*, 57–64.
- Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science*, *314*, 1560–1563.
- Nowak, M. A., May, R. M., & Sigmund, K. (1995, June). The arithmetics of mutual help. *Scientific American*, *272*, 50–55.
- Nowak, M. A., & Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. *Nature*, *393*, 573–577.
- Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, *437*, 1291–1298.
- Oda, R. (1997). Biased face recognition in the prisoner's dilemma games. *Evolution and Human Behavior*, *18*, 309–315.
- Olson, M. (1965). *The logic of collective action: Public goods and the theory of groups*. Cambridge, MA: Harvard University Press.
- Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge, U.K.: Cambridge University Press.
- Ostrom, E. (2003). Toward a behavioral theory linking trust, reciprocity and reputation. In E. Ostrom & J. Walker (Eds.), *Trust and reciprocity: Interdisciplinary lessons from experimental research* (pp. 19–79). New York: Sage.
- Ostrom, E., Walker, J., & Gardner, R. (1992). Covenants with and without a sword: Self governance is possible. *American Political Science Review*, *86*, 404–417.
- Page, T., Putterman, L., & Unel, B. (2005). Voluntary association in public goods experiments: Reciprocity, mimicry and efficiency. *The Economic Journal*, *115*, 1032–1053.
- Panchanathan, K., & Boyd, R. (2003). A tale of two defectors: The importance of standing for the evolution of indirect reciprocity. *Journal of Theoretical Biology*, *224*, 115–126.
- Panchanathan, K., & Boyd, R. (2004). Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature*, *432*, 499–502.
- Price, M. E. (2003). Pro-community altruism and social status in a Shuar village. *Human Nature*, *14*, 191–208.
- Price, M. E. (2006). Monitoring, reputation, and ‘greenbeard’ reciprocity in a shuar work team. *Journal of Organizational Behavior*, *27*, 201–219.
- Price, M. E., Cosmides, L., & Tooby, J. (2002). Punitive sentiment as an anti-free rider psychological device. *Evolution and Human Behavior*, *23*, 203–231.
- Queller, D. C., Ponte, E., Bozzaro, S., & Strassmann, J. E. (2003). Single-gene greenbeard effects in the social amoeba *dictyostelium discoideum*. *Science*, *299*, 105–106.
- Richerson, P. J., & Boyd, R. (2004). *Not by genes alone: How culture transformed human evolution*. Chicago: University of Chicago Press.

- Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S., & Kilts, C. D. (2002). A neural basis for social cooperation. *Neuron*, *35*, 395–405.
- Sheldon, K. M., Sheldon, M. S., & Osbaldiston, R. (2000). Prosocial values and group-assortation within an n-person prisoner's dilemma. *Human Nature*, *11*, 387–404.
- Sinervo, B., Chaine, A., Clobert, J., Calsbeek, R., Hazard, L., Lancaster, L., et al. (2006). Self-recognition, color signals, and cycles of greenbeard mutualism and altruism. *Proceedings of the National Academy of Sciences*, *103*, 7372–7377.
- Singer, T., Kiebel, S. J., Winston, J. S., Dolan, R. J., & Frith, C. D. (2004). Brain responses to the acquired moral status of faces. *Neuron*, *41*, 653–662.
- Sober, E., & Wilson, D. S. (1998). *Unto others: The evolution and psychology of unselfish behaviour*. Cambridge, MA: Harvard University Press.
- Stevens, J. R., & Hauser, M. D. (2004). Why be nice? Psychological constraints on the evolution of cooperation. *Trends in Cognitive Sciences*, *8*, 60–65.
- Stone, V. E., Cosmides, L., Tooby, J., Kroll, N., & Knight, R. T. (2002). Selective impairment of reasoning about social exchange in a patient with bilateral limbic system damage. *Proceedings of the National Academy of Sciences*, *99*, 11531–11536.
- Sugiyama, L. S., Tooby, J. T., & Cosmides, L. (2002). Cross-cultural evidence of cognitive adaptations for social exchange among the Shiwiar of Ecuadorian Amazonia. *Proceedings of the National Academy of Sciences*, *99*, 11537–11542.
- Takezawa, M., & Price, M. E. (2006, September). *Revisiting "the evolution of reciprocity in sizable groups."* Paper presented at the 4th CEFOM Symposium 21 International "Cultural and Adaptive Bases of Human Sociality," International House of Japan, Tokyo.
- Tinbergen, N. (1968). On war and peace in animals and man: An ethologist's approach to the biology of aggression. *Science*, *160*, 1411–1418.
- Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, *46*, 35–57.
- Trivers, R. L. (2004). Mutual benefits at all levels of life. *Science*, *304*, 964–965.
- Trivers, R. L. (2006). Reciprocal altruism: 30 years later. In P. M. Kappeler & C. P. van Schaik (Eds.), *Cooperation in primates and humans: Mechanisms and evolution*. New York: Springer.
- Vogel, G. (2004). The evolution of the golden rule. *Science*, *303*, 1128–1131.
- Wedekind, C., & Milinski, M. (2000). Cooperation through image scoring in humans. *Science*, *288*, 850–852.
- Williams, G. C. (1966). *Adaptation and natural selection*. Princeton, NJ: Princeton University Press.
- Williams, G. C. (1992). *Natural selection: Domains, levels and challenges*. Oxford, U.K.: Oxford University Press.
- Williamson, O. E. (1996). *The mechanisms of governance*. New York: Oxford University Press.
- Wilson, D. S. (2001). Evolutionary biology: Struggling to escape exclusively individual selection. *Quarterly Review of Biology*, *76*(2), 199–205.
- Wilson, D. S. (2002). *Darwin's cathedral: Evolution, religion, and the nature of society*. Chicago: University of Chicago Press.
- Wilson, D. S. (2006). Human groups as adaptive units: Toward a permanent consensus. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind: Culture and cognition*. (pp. 78–90). Oxford, U.K.: Oxford University Press.
- Wilson, D. S., & Sober, E. (1994). Reintroducing group selection to the human behavioural sciences. *Behavioral and Brain Sciences*, *17*(4), 585–654.
- Wilson, E. O. (1975). *Sociobiology: The new synthesis*. Cambridge, MA: Belknap Press.
- Wilson, E. O. (2000). *Sociobiology: The new synthesis, Twenty-fifth anniversary addition*. Cambridge, MA: Belknap Press.
- Yamagishi, T., Tanida, S., Mashima, R., Shimoma, E., & Kanazawa, S. (2003). You can judge a book by its cover: Evidence that cheaters may look different from cooperators. *Evolution and Human Behavior*, *24*, 290–301.
- Zahavi, A. (1975). Mate selection—A selection for a handicap. *Journal of Theoretical Biology*, *53*, 205–214.
- Zahavi, A. (1995). Altruism as handicap: The limitations of kin selection and reciprocity. *Journal of Avian Biology*, *26*, 1–3.