

# Chapter 7

## The Logic and Location of Strong Reciprocity: Anthropological and Philosophical Considerations

Jordan Kiper and Richard Sosis

### Introduction

Many behavioral economists and evolutionary anthropologists claim that a cornerstone of human cooperation is the willingness to pay the costs of helping cooperators or punishing cheaters, which is known as *strong reciprocity* (e.g., Bowles, Boyd, Matthew, & Richerson, 2012; Boyd & Richerson, 2009; Diekmann, Jann, Przepiorka, & Wehrl, 2014; Gintis, Henrich, Bowles, Boyd, & Fehr, 2008; Gintis & Fehr, 2012). Yet critics have put forth several reasons purporting to challenge the very idea of strong reciprocity (e.g., Burnham & Johnson, 2005; Guala, 2012; Hagen & Hammerstein, 2006; Price, 2008; Yamagishi et al., 2012). In this chapter, we examine some of these criticisms and related challenges through anthropological and philosophical lenses, and provide a few ethnographic examples of wartime altruism to illustrate the difficulties of isolating strong reciprocity in the real world.

Semantic issues should be acknowledged here at the outset, including the fact that the term *strong reciprocity* is not a straightforward one. Hearing the words *strong* and *reciprocity* together, one would assume that what was being discussed was the robust exchange of something for mutual benefit—but that is not entirely the case. The term *strong reciprocity* designates one of two things that both entail a cost for an agent: rewarding cooperators when it would be more advantageous to exploit them or punishing defectors when it would be more advantageous to ignore them (Gintis, 2000a, p. 177). In either case, the key is that an agent pays a high price for enforcing reciprocity among others, but does so without any personal benefit for himself or herself, which defies traditional models of self-interested maximization in economics and biology (Gintis et al., 2008, p. 243). Evolutionary theory has thus crept into the picture to answer the question as to *why* anyone would behave in such a manner.

---

J. Kiper (✉) • R. Sosis

Department of Anthropology, University of Connecticut, Storrs, CT, USA  
e-mail: [jordan.kiper@uconn.edu](mailto:jordan.kiper@uconn.edu); [richard.sosis@uconn.edu](mailto:richard.sosis@uconn.edu)

© Springer International Publishing AG 2017

M. Li, D.P. Tracer (eds.), *Interdisciplinary Perspectives on Fairness, Equity, and Justice*, DOI 10.1007/978-3-319-58993-0\_7

107

There is nevertheless an equally important issue that has come to vex those who defend strong reciprocity. Does the behavior even exist outside of laboratory experiments? Metaphysics aside, the question is empirically motivated, insofar as evidence for strong reciprocity comes almost entirely from cross-cultural studies of economic games. Moreover, field research has centered on the costs and benefits of individual third-party punishments, which are rare and notoriously difficult to measure. So three questions persist: whether strong reciprocity is an artifact of economic games, whether it occurs in the real world, and, if so, why did it evolve?

A general consensus among critics and defenders is that these queries cannot be fully answered (or dismissed) without more data and, most importantly, a unified evolutionary theory of justice (see Debove, Baumard, & Andre, 2016). The result is that strong reciprocity remains an active and dynamic area of research in economics, psychology, and anthropology. Our aim here is to advance this line of research by approaching strong reciprocity from two different perspectives and thereby making two specific contributions. First, we take a philosophical stance and outline the logical argument for strong reciprocity in detail, drawing attention to its most questionable premises. Second, we take an anthropological approach and address what we see as the most critical issue facing strong reciprocity, which is that there is little evidence for strongly reciprocal behavior in the real world, outside of economic games. We conclude that (1) despite some weak premises, the foundational argument for strong reciprocity is logically sound, and (2) while it is very unlikely that strong reciprocity is an artifact entirely limited to experimental settings, it is difficult to detect the behavior in nonexperimental contexts. Lastly, we suggest that while the impulses of strong reciprocity can motivate justice and fairness, one of the reasons that strong reciprocity is difficult to detect in real-world contexts is that cultural forces influence and often limit the manifestation of strong reciprocity impulses.

## Strong Reciprocity

Ever since Herbert Gintis' publication "Strong Reciprocity and Human Sociality" (2000a), economists and evolutionary biologists have broadly classified reciprocity as either weak or strong. *Weak reciprocity* is tit-for-tat behavior that benefits, or is at least optimal, for reciprocating agents, while *strong reciprocity* is cooperative behavior that is suboptimal for the practicing agent (Guala, 2012, p. 1). Broadly speaking, weak reciprocity operates efficiently in societal contexts or cultures where there are visible credentials for agents, such as image-scoring or reputational score keeping, which concerns someone's perceived quality. Strong reciprocity, on the other hand, is expected to occur in societal contexts where the previously mentioned credentials are absent, as in large societies where there is an immense variance in the likelihood of iterated cooperation. What makes strong reciprocity so remarkable is that it is a selfless policing behavior insofar as an agent freely rewards or punishes others at a personal cost.

## *Cooperation and Justice*

Besides diverging from rational choice theory, strong reciprocity touches upon two major topics in the behavioral and brain sciences. The first is *cooperation*, which here means any process by which individuals or groups coordinate their actions for mutual benefit (Axelrod, 1984, p. 6). This consists of kin selection (Hamilton, 1964) and altruistic behavior such as direct, indirect, or network reciprocity (see Alexander, 1987; Nowak & Sigmund, 1998; Trivers, 1971); adaptive behaviors such as costly signaling or self-imposed handicaps (Sosis, 2006; Zahavi, 1975); and coaptations of language, communication, and social cognition for coordinating group efforts (e.g., Moll & Tomasello, 2007). The second topic is *justice*, which is understood widely enough to include the human proclivity for fairness when exchanging resources, enjoying privileges, and enforcing punishments (Rawls, 1971, pp. 8–9). Fairness consists of comparing one's efforts and subsequent rewards with those of others as well as caring about equity (e.g., Brauer & Hanus, 2012). Because doing so allows one to detect cheaters or persons whose rewards are greater than their efforts, justice goes hand in hand with fairness such that justice itself is thought of as fairness (see Rawls, 1971).

Of course, justice and fairness also share a close relationship with cooperation. Without fairness and reciprocity, the mutual trust between individuals is severed and the coordinated efforts of groups collapse, resulting in overall lost benefits and decreased fitness compared to cooperative groups (Axelrod & Hamilton, 1984). This in turn raises the question about the proximate mechanisms that bring about justice. Under this topic have come numerous anthropological accounts about various reciprocal behaviors that maintain social bonds (e.g., Mauss, 1990/1950; Sahlins, 1972) and psychological descriptions of communicative strategies that influence social exchanges (e.g., Cialdini, 2006). But only over the last decade have neuroscientists shown that justice is rooted in what is best described as *moral emotion*. Whenever we help someone in need, our reward centers are activated, including the subgenual region, which is associated with oxytocin and social attachment. The result is that when we help, we often experience a “warm glow”—a feeling of pleasure in doing good—that constitutes an emotional basis for engaging in moral acts, thus accounting for many costly behaviors (e.g., Andreoni, 1990). Similarly, when witnessing unfairness, we experience negative emotions and action patterns generated by neural substructures such as the anterior insula (Hsu, Anen, & Quartz, 2008; Kaltwasser, Hildebrandt, Wilhelm, & Sommer, 2016; Knoch et al., 2008; Tabibnia, Satpute, & Lieberman, 2008). This discovery identifies a cognitive mechanism for justice and with it a rather unexpected result. Rather than responding only when we alone experience injustice, our moral emotions are triggered whenever we see anyone experiencing an injustice, including strangers (e.g., Mendez, 2009; see also Sanfey, Rilling, Aaronson, Nystom, & Cohen, 2003).

It is here that strong reciprocity enters the picture. In lab experiments where individuals witness one participant cheating another, there is heightened activity in the anterior insula. Yet in experiments where individuals can actually punish the

cheater, they also experience activity in the caudate nucleus, a brain region dedicated to learning, reward, and pleasure (de Quervain et al., 2004; Luo et al., 2006; Pascual, Rodrigues, & Gallardo-Pujol, 2013). Similar brain regions are activated whenever individuals see a participant cooperating with others and seek to reward them for doing so (Li & Yamagishi, 2014; Watanabe et al., 2014). Remarkably, individuals in many laboratory experiments will go out of their way—even giving up their own resources—to punish cheaters and reward cooperators (e.g., Engel, 2011; Fehr & Fischbacher, 2004; Fehr & Gächter, 2002). These data have led researchers to label such behavior as *strong reciprocity* and to speculate about its ultimate cause.

### ***Laboratory Experiments***

Most of the evidence for strong reciprocity comes from experiments involving economic games such as the dictator, ultimatum, and public goods game (see Fehr & Gächter, 2001, 2002; Fehr & Fischbacher, 2004; Fischbacher, Gächter, & Fehr, 2001). In each of these games, participants are given money and rules for playing out a game simultaneously and anonymously with other players in a lab, usually over a computer interface. Because participants can increase their earnings, it is expected that players will adopt a rational strategy in which they pursue ordered preferences to maximize self-interest, which is presumed to be money earned during the game itself. Under most circumstances, however, participants typically do not maximize their earnings but rather the perceived equity among players.

As a brief sketch, consider the nature of the ultimatum game, which involves the interplay of anonymous and unseen participants. At the onset of play, participant P receives an amount of money to offer participant S, who is usually located in another room. If S accepts P's offer, P keeps the remainder, but if S rejects the offer, both P and S get nothing (Henrich, Boyd et al., 2005). Because the game allows its participants to behave selfishly, it is expected that P will offer as little money as possible to S. Likewise, S is expected to accept whatever P offers, since any offer gives S more than he or she possesses. But participants tend to split their resources (Fehr & Fischbacher, 2003). This is unexpected in light of the neoclassical economic view known as *Homo economicus* or “human the self-interested.” In other words, the game is one-shot and the participants remain anonymous, so there is no immediate or long-term reward for P to benefit S or vice versa—and yet most participants forsake self-maximization to benefit others.

The behavior not only challenges paradigmatic views in neoclassical economics but also traditional evolutionary theories of cooperation. Because strong reciprocity takes place between unrelated individuals and does not contribute to inclusive fitness, it cannot be explained by kin selection theory. Since it occurs in economic experiments that involve one-shot interactions in which participants cannot later reciprocate, it cannot be explained by the theory of reciprocal altruism either. Similarly, because participants are anonymous and thus cannot earn a reputation, it

cannot be explained as a form of indirect reciprocity. Finally, it is unlikely that strong reciprocity is a costly signal or handicap indicating the participant's type-quality because there is no subsequent interaction between participants (see Fehr, Fischbacher, & Gächter, 2002, p. 10).

### *Field Experiments*

Such anomalousness raises the question of whether strong reciprocity occurs outside the laboratory. For those who defend strong reciprocity, there are several reasons for believing that it does. First, humans are often willing to aid strangers when an audience is absent and reciprocity is unlikely—a point we shall return to later when discussing persons disrupted by warfare (Gintis et al., 2008, p. 251). Second, humans exercise costly social-norms and conform to social expectations even when they are alone and unobserved or among complete strangers (Bowles & Gintis, 2002, p. 125). Third, the connection between strongly reciprocal behavior and moral emotions for fairness suggests that prosociality and notions of justice are indeed motivated by strong reciprocity (Gintis, Bowles, Boyd, & Fehr, 2003, p. 154).

Critics of strong reciprocity nonetheless find these examples unconvincing. The problem is that it is often difficult to determine whether a behavior is an instance of strong reciprocity or another evolved form of cooperation. For instance, aiding strangers, conforming to social norms, and entertaining moral notions of justice can be just as easily rationalized by inclusive fitness (West, Mouden, & Gardner, 2011, p. 252). In light of such criticisms, ethnographic field experiments involving variations of the previously mentioned games have been carried out in non-western countries (e.g., Fehr & Leibbrandt, 2011; Rustagi, Engel, & Kosfeld, 2010). While these cross-cultural experiments reveal that every sampled culture thus far exemplifies some degree of strongly reciprocal behavior, they also show that the behavior varies according to the participant's cultural understanding of reciprocity and economic interactions.

For example, in Papua New Guinea, participants in the ultimatum game reject any offer greater than 50% (see Tracer, 2003). The reason for this variation is that relative to other cultures, the people of Papua New Guinea give modest gifts (i.e., equivalent to a 10% offer in economic games) to signal affection between kin and allies, but large gifts (i.e., roughly equivalent to or greater than a 50% offer) to ingratiate recipients. Ingratiation in Papua New Guinea is known to engender long-term servitude between partners, which is in the direction of a receiver serving the allocator, as it reflects the traditional mode of economic exchange and tribal politics. Accepting large gifts is thus avoided, even in experiments involving economic games. This is not the case for westerners or anyone else steeped in a market economy, where splitting shares (i.e., equivalent to a 50% offer in economic games) is a sign of mutualism, which facilitates economic partnerships in market interactions (Henrich & Boyd, 2001).

As a result, the defenders of strong reciprocity are correct when they observe that experiments involving economic games cross-culturally elicit strongly reciprocal behavior. Yet the cultural variation in strong reciprocity raises questions about its ontogeny and enculturation. These include questions about the inculcation of reciprocal norms during childhood (Feldman, 2015), the internalization of one's cultural norms regarding economic exchange (e.g., Guiso, Sapienza, & Zingales, 2009), and whether the behavior is simply an artifact of economic games (Guala, 2012; Yamagishi et al., 2012).

## The Evolution of Cooperation

Despite its alleged limitation to economic games, strong reciprocity is said to be central to the evolution of human cooperation (e.g., Bowles et al., 2012; Gintis & Fehr, 2012). This may seem counterintuitive if one reflects on the principle of self-interest in traditional economics or evolutionary biology. In both cases, it is presumed to be in the best interest of agents to be self-maximizing—in fact, doing so generally leads to Nash equilibria (i.e., where agents gain nothing by unilaterally changing their behavior if they know the strategies of other agents). Economically speaking, this means increasing one's own profits and, in biological terms, maximizing one's inclusive fitness (e.g., Bowles & Gintis, 2002). Hence, an adaptive strategy for self-interested agents is weak reciprocity: to cooperate directly or indirectly with relatives or known reciprocators and to avoid cheaters.

However, this picture of human cooperation is incomplete if one reflects on what is required for weak reciprocity to function among groups. Such a group would need to have a history of interaction and the potential for future interactions; be relatively small or at least not excessively large and anonymous; and freely circulate reputational information about individuals among the group (Guala, 2012). Of course, these conditions do not always hold in human communities such as state-level societies where an individual regularly interacts with strangers. Likewise, it is not unusual for persons to act selflessly toward strangers or to give anonymously, and to do so without expecting direct or indirect reciprocity, but rather, as many philanthropists say, “because it is the right thing to do.” Strong reciprocity is said to fill these gaps: it occurs outside the conditions of weak reciprocity and accounts for ostensibly selfless behavior (see Fehr et al., 2002). Moreover, having strong reciprocators in a group is potentially adaptive insofar as they support cooperation by rewarding cooperators, and, most importantly, enforce cooperation by engaging in the costly punishment of cheaters (e.g., Bowles & Gintis, 2004; Boyd, Gintis, Bowles, & Richerson, 2003; Gintis, 2000a).

These claims have generated controversy. First, it is not clear whether enforcing cooperation by punishing cheaters (*negative strong reciprocity*) occurs as much in the real world as selflessly supporting cooperators (*positive strong reciprocity*). Second, it is difficult to see how strong reciprocity would be adaptive when strong reciprocators are likely to incur diminished gains compared to weak reciprocators (Guala, 2012). In what follows, we address the issue of diminished gains and return to negative versus positive reciprocity.

## ***Group Selection Theory***

The problem of diminished gains is obviated by group selection theory or the idea that nature can select at the level of groups (e.g., Sober & Wilson, 1998; Wilson & Sober, 1994). Albeit a somewhat complex topic, which often gets muddled or glossed over by critics of strong reciprocity (see Pisor & Fessler, 2012), group selection theory can be understood as follows. If groups conform to different behaviors, then differences minimize within those groups but maximize between them. When faced with threats, such variability allows some groups to be more successful than others and thereby be more adaptive (Leland & Brown, 2002, p. 64). Of course, this is not to say that members within cooperative groups are in any way equal or that group selection benefits every individual, since what most likely contributes to group selection is a shift in prosocial sentiments that favor central or powerful group members (e.g., Baldassarri, 2013). Hence, it is most likely that group selection is akin to reframing perceived equity, such that cooperative groups outcompete less cooperative ones.

Championing this view, evolutionists interested in strong reciprocity have argued that strongly reciprocal behavior is costly for individuals but adaptive for groups (e.g., Boyd, Gintis, & Bowles, 2010; Boyd et al., 2003; Gintis, 2000b). Specifically, it would be especially adaptive when populations become large and anonymous or when the *shadow of the future* (i.e., anticipation of future reciprocal interactions between individuals) is cut short by culturally disruptive phenomena such as natural disasters or warfare. In these circumstances, strong reciprocators would enforce group cooperation while purely weak reciprocators would not (Fehr & Fischbacher, 2003, p. 790). Over time groups with strong reciprocators would fare better than those with only weak reciprocators, eventually allowing the former to outcompete, overtake, or absorb the latter (see Fehr et al., 2002; Henrich & Boyd, 2001). Because these circumstances characterize most cultures since the Neolithic, they entail that strong reciprocity would have been an adaptive behavior and that group selection would serve as the mechanism for stabilizing it across human populations (Gintis et al., 2008, p. 241).

But isn't this argument in conflict with traditional evolutionary biology? Similar to defenders of strong reciprocity today, V.C. Wynne-Edwards (1962, 1964) once argued that organisms cooperate for the welfare of their species, to which George Williams (1966) famously replied that cooperation is just like any other behavior: it is fully explicable at the level of genes and a fortiori the fitness of the individual. For most of the twentieth century, developments in evolutionary biology were on Williams's side (e.g., Dawkins, 2006/1976). It was widely believed that because genes are the heritable element behind selected phenotypes, the individual is in fact the level at which natural selection occurs. Hence, there was no need to resort to the group level when accounting for naturally selected behavior.

Nonetheless, evolutionists toward the end of the twentieth century and early twenty-first century began recognizing two things. First, terms such as *inclusive fitness*, *kin selection*, and *group selection* were not mutually exclusive terms or

competitive explanations of adaptiveness, as they had so often been characterized. Instead, they were different ways of discussing the same thing, namely, the manner in which nature works at multiple levels of phenomena when selecting adaptive traits (see Nowak, Tarnita, & Wilson, 2010). Second, and related to the prior observation, groups could indeed stand as adaptive units (Wilson & Sober, 1994). Lastly, strong reciprocity is observed in nature among bacteria and other organisms (Inglis, West, & Buckling, 2014).

To illustrate, consider a classic thought experiment by John Maynard Smith (1964), which was ironically designed with the intent to counter group selection but actually highlights its logic. If we imagine two haystacks sitting side by side and containing mice, and if those mice gathered resources from the environment just outside their haystacks but mated only with mice from within their own haystack, then they would experience two levels of selection. One would exist between individuals within their own haystack (*individual selection*) and the other between the two populations (*group selection*). For example, if one population gathered more resources than the other, the more resourceful haystack would have a competitive advantage over the other, such as surviving a harsh winter, leaving more offspring, and thus increasing their fitness. Over time nature would favor the alleles of mice from the more resourceful haystack.

Another way of saying this is that there is multilevel selection. Genes are nested within cells, which are nested within organisms, who are themselves nested within groups. The survival of any trait is the effect of nature selecting at the levels of groups all the way down to genes and genetic drift (Grafen, 1985). Therefore, whenever a group trait is selected, so too are underlying genes within individuals of the group (Wilson & Sober, 1994).

## ***Gene-Culture Coevolution***

Defenders of strong reciprocity have gone one step further. According to *dual-inheritance theory* or *gene-culture coevolution*, genes engender humans capable of culture, and culture is effectively the construction of a niche that in turn creates pressures selecting for certain genes (Gintis et al., 2008, p. 247). In other words, while nature can act on groups and therein select for individual traits, human groups create culture and culture can engender additional selective pressures on individuals within the group. This dynamic is especially significant for humans such that it is responsible for numerous species characteristics. For instance, the social advent of herding brought about selective pressures that favored human genes that extended the ability to digest lactose beyond early childhood, which endowed groups with preferences for milk and this in turn compelled them to transform their natural environment to facilitate that preference (Leland & Brown, 2002). Numerous other examples could be given, including the advent of writing and cultural transformations due to technology and science (Cochran & Harpending, 2009). The point is



that the human genome allows individuals to transform their natural environment so as to facilitate social arrangements. Moreover, these arrangements create a niche that constrains and promotes aspects of the human genome, thus selecting for patterns of cognition, affection, and behavior.

It is theorized that strong reciprocity emerged from a process of gene-culture coevolution. According to Gintis (2011), what got the whole process going was the selection for phenotypic plasticity in dynamically changing ancestral environments. With phenotypic plasticity came the capacity to learn and thus the epigenetic transmission of information otherwise known as culture. Having the capacity for learning and communicating cultural innovations to subsequent generations, early human communities developed norms supporting weak reciprocity (p. 881). This would explain the selective pressures for an accompanying set of prosocial traits that appear to have emerged in early human communities such as moral indignation, guilt, and empathy (Sterelny, 2011). Such traits are rooted in nonhuman primates, including old-world monkeys, who also experience empathy and moral emotions (Dugatkin, 1999). These in turn would have generated moral values and the internalization of prosocial norms to induce community members into conforming to social duties (Gintis, 2011, p. 881). With the advent of cultural technologies for internalizing social norms, such as religion, culture would have put additional selective pressures on neural structures for prosociality. As numerous ethnographic studies show (e.g., Cushing, 1998; Grusec & Kuczynski, 1997; Nisbett & Cohen, 1996), a distinguishing feature of internalizing norms is that individuals are taught—sometimes with great intensity as with rites of passage—to behave prosocially even when community members are not observing them. With such technologies neural structures for internalizing and practicing social norms would have then been privileged in human evolution (Gintis, 2011, p. 881).

The tendency for strong reciprocity would thus emerge from neural structures dedicated to weak reciprocity and prosociality such as the superior temporal sulcus (Moll et al., 2005), interior insula (e.g., Hsu et al., 2008), and caudate nucleus (e.g., Pascual et al., 2013). However, these could then be co-opted to respond to more wide-ranging forms of altruism, such as expressing more indignation to injustices outside of one's kin or affines, by selective pressures at the group level. Indeed, it is possible that strong reciprocity is related to human niche specialization, such that it emerged out of social conflict as an alternative social option, which resulted in less conflict and reduced social problems (Bergmuller & Taboorsky, 2010). The argument here is a familiar one for any group selected trait. When early human communities acquired strong reciprocators, they cooperated more than communities with only weak reciprocators, which brought about selective pressures that favored alleles for the neural substrates underlying strong reciprocity (Gintis, 2003, p. 407). The selection of these genetic factors most likely “ratcheted” the behavior, increasing strong reciprocity and allowing groups of strong reciprocators to outcompete less cooperative groups or even drive them into extinction. This scenario likely began early in human evolution but was enhanced with the appearance of settled communities around 10,000 years ago (e.g., Boyd & Richerson, 2009).

## The Logic of Strong Reciprocity

We are now in a position to appreciate the overall logic of strong reciprocity as a scientific idea and thereby see exactly where critics take aim. Here is a thumbnail sketch of the argument for strong reciprocity that is based on its chief premises (P1–10) as discussed in several articles (Bowles et al., 2012; Bowles & Gintis, 2002; Boyd et al., 2010; Boyd et al., 2003; Fehr & Fischbacher, 2003; Fischbacher, Gächter, & Fehr, 2001; Fehr & Henrich, 2003; Fehr & Rockenbach, 2003; Schneider & Fehr, 2010; Gintis, 2000a, 2000b, 2011; Gintis et al., 2003; Gintis & Fehr, 2012; Gintis et al., 2008; Henrich et al., 2006).

P1	Strong reciprocity appears to be a type of altruistic behavior.
2	Altruistic behaviors are attributable to some predisposition to cooperate with others.
3	The predisposition to cooperate with others reduces to self-interest.
4	Strong reciprocity must reduce to self-interest (from 1, 3).
5	But laboratory and field experiments indicate that strong reciprocity is not motivated by self-interest.
6	Some altruistic behaviors do not reduce to self-interest (from 2, 5).
7	It is possible that strong reciprocity results from group selection.
8	Strong reciprocity can sustain cooperation in the face of group threats.
9	If strong reciprocity can sustain groups, then it is adaptive when groups face famine, war, or dispersal—all of which were prevalent during human evolution.
10	Strong reciprocity is possibly adaptive (from 8, 9).

This can be spelled out a bit further as follows.

Based on behavioral economic experiments, (1) strong reciprocity is a distinct type of altruistic behavior insofar as it is detrimental to the agent performing it but beneficial to another. For instance, “rejections in the ultimatum game can be viewed as altruistic acts because most people view the equal split as the fair outcome” (Fehr & Fischbacher, 2003, p. 786). (2) Altruistic behaviors are cooperative behaviors insofar as they are conducive to reciprocity. However, (3) when an organism cooperates with others, it does so in virtue of some naturally selected predisposition such that (4) any predisposition for cooperation must derive from the organism’s self-interest for maximizing resources or inclusive fitness (e.g., Gintis et al., 2008; Henrich et al., 2004). At least that much seems clear with regard to neoclassical economic theory and traditional evolutionary biology. (5) But “humans often cooperate in ‘one-shot’ interactions” and “in these situations there is little chance of direct or indirect reciprocation, so self-interest-based explanations of cooperation are unconvincing” (Bowles & Gintis, 2002, p. 125). (6) Because kin selection theory, reciprocal altruism, indirect reciprocity, and costly signaling cannot account for strong reciprocity, it does not reduce to traditional theories of self-interest (e.g., Fehr et al., 2002, p. 10). (7) If strong reciprocity sustains group cooperation, then it is selected at the group level. (8) Recent theoretical models of gene-culture coevolution show that strong reciprocity is capable of generating within-group cooperation

where weak reciprocity would not, and thus giving such groups an advantage over others (e.g., Fehr & Fischbacher, 2003, p. 790). (9) Throughout human evolution, groups faced extreme threats of famine, war, and dispersal. Groups with strong reciprocators would have survived these threats where purely weak reciprocators would not because strong reciprocity reinforces cooperation (see Fehr, Fischbacher, & Gächter, 2001; Fehr & Henrich, 2003; Henrich & Boyd, 2001). Therefore, data on strong reciprocity and gene-culture coevolution suggest that strong reciprocity is an adaptive behavior, which was unrecognized in science until experiments revealed its importance (e.g., Gintis, 2011).

## Criticisms and Potential Challenges

Taking stock of the above argument, it is clear to us that the basic logic is valid. The main question then is whether it is also sound. In this section, we highlight three criticisms that draw into question some of the premises behind the argument for strong reciprocity and thereby point to issues that require further empirical and theoretical investigation.

### *Evidence in “the Wild”*

The perennial challenge put forth by critics takes aim at the first premise and its underlying assumption that experimental data sufficiently demonstrate that strong reciprocity is a behavior in the real world. Critics argue that the ethnographic data for strong reciprocity, which allegedly demonstrate the behavior “in the wild,” are simply cross-cultural economic experiments that replicate the very conditions in which the behavior was originally identified (e.g., Price, 2008; Trivers, 2006). Responding to this criticism, defenders of strong reciprocity have cited several ethnographic studies purporting to describe altruistic punishment and thus various examples of negative strong reciprocity (e.g., Henrich et al., 2004; Marlowe et al., 2008). However, critics point out that these studies can be interpreted in numerous ways and that even the original ethnographers who recorded them are unsure as to whether the punishments they observed constitute strong reciprocity. In short, costly punishment observed in ethnographic settings is usually described as collective retribution or coalitional punishment, designed as such to offset the costs of punishing free riders and, thus, obviating the risk of negative strong reciprocity (e.g., Boehm, 2012).

Furthermore, because punishments observed in ethnographic settings are almost always balanced reciprocity between individuals or collective third-party punishment, it is difficult to confidently identify such behaviors as strong reciprocity. The gap between ethnographic and experimental evidence has led many critics to claim

that strong reciprocity is an artifact of economic games (see Guala, 2012; Hsu et al., 2008). We consider this criticism to be the central problem to strong reciprocity and one we shall address throughout the rest of this chapter.

For now, we wish to stress that several researchers of strong reciprocity have responded that critics adopt an understanding of experimental data that is too narrow, and that a wider interpretation is not only valid but also more fruitful (Bowles et al., 2012; Gintis & Fehr, 2012; Henrich & Chudek, 2012). A “narrow” interpretation of strong reciprocity is that behavior in economic games is invaluable for shedding light on the proximate psychological motives and enculturated reactions to violations of social norms. Beyond that, any claim that strong reciprocity is an evolved behavior imports more than what is warranted by the data. A “wide” interpretation is that experiments involving economic games simplify the conditions of cooperation in the real world and isolate the costs of strong reciprocity that are difficult to measure in ethnographic settings (see Guala, 2012, p. 5). Moreover, these experiments are internally valid insofar as they correctly identify the proximate mechanisms of strong reciprocity and are externally valid insofar as they help rationalize strong reciprocity in the real world (e.g., Bowles et al., 2012). Although the latter claim is contested, it is worth stressing that the external validity of any experiment is conjectural and that the conjectures made by defenders of strong reciprocity are well grounded.

Several experimenters have shown, for instance, that strongly reciprocal behavior in laboratory settings significantly correlates with behavior observed in various field experiments (e.g., Henrich, Heine, & Norenzayan, 2010). These experiments reveal the bare costs that persons are willing to pay in order to sustain cooperation, and this helps shed light on the ways in which cultures use proclivities for justice and cooperation to collectively control for freeriding while minimizing retaliatory costs against strongly reciprocal individuals. Experiments may also reveal strategies for human cooperation that are expressed differently in the real world. Consider the example of ostracism. In economic experiments, punishment is rendered by direct ostracism or ending all cooperation with a defector, which is costly to the punisher. However, this is rarely observed in ethnographic settings, most likely because it is easier for humans simply to avoid defectors, which is costly but not as drastic as laboratory behavior. Finally, group selection theory provides a theoretical framework to explain the ultimate cause of the behavior and to rationalize the ubiquity of strong reciprocity in various cross-cultural field experiments as well as neurological studies of injustice and cooperation (see Pisor & Fessler, 2012).

### *Type Distinction and Adaptiveness*

Nonetheless, the issue of lacking concrete evidence for strong reciprocity outside of laboratory experiments provides the grounds for additional criticisms. One is that without further real-world evidence, it is still possible to question what strong reciprocity is exactly (Price, 2008). The argument is that the nature of economic experiments is

purposefully restricting—for example, limiting participants to an anonymous interaction that is often one-shot, which is done, of course, to isolate variables of interest. However, doing so challenges external validity in the case of strong reciprocity, because the behavior was identified within experimental settings and it has been difficult to document outside of such settings. As a result, strong reciprocity could be the basic impulse for reciprocity as it gets expressed in unusual settings such as the ultimatum game (Trivers, 2006, p. 965).

Another criticism centers on the premise that strong reciprocity is adaptive. Critics note that the argument for strong reciprocity pivots between experimental evidence and real-world behavior (Burnham & Johnson, 2005). Specifically, defenders use real-world behaviors, such as weak reciprocity, to make sense of strong reciprocity in experimental settings. They argue that the behavior is a distinct one, because, in economic settings, it cannot be reduced to kin selection, direct reciprocity, indirect reciprocity, or costly signaling (Fehr et al., 2002, p. 10). Using this mode of reasoning, critics argue that if strong reciprocity is adaptive, then its adaptiveness must also apply to experimental settings. After all, defenders argue strongly that strong reciprocity is adaptive because of its group-level benefits. But in the context of economic games, strong reciprocity is conferred to unknown persons and not to the agent's group. Hence, the behavior is not adaptive (Burnham & Johnson, 2005, p. 122).

As a defense, it should be stressed that this argument is somewhat of a modal mischaracterization, that is, an inaccurate portrayal of the purported truth conditions of premises seven and ten from above. For defenders of strong reciprocity, the behavior is not necessarily adaptive in particular economic games but rather possibly adaptive outside of games at the group level, which is a reasonable proposition given the consistent emergence of the behavior in experimental settings, despite the difficulty of detecting it in the real world (see Fehr & Fischbacher, 2003; Fehr & Henrich, 2003; Henrich & Boyd, 2001). Nevertheless, critics also argue that even if we assume strong reciprocity is adaptive for groups, the infrastructures of economic games do not apply to real-world settings. For instance, one-shot anonymous interactions between individuals from different groups in the real world would still be acting in such a way that the behavior would not benefit any group (Burnham & Johnson, 2005, p. 122).

Even though this criticism could be dismissed for confusing the modal propositions comprising the argument of strong reciprocity, it highlights a peculiarity about strong reciprocity that once again arises from experimental evidence and its application to the real world. Defenders of strong reciprocity argue that the behavior observed in economic games can be applied to real-world behavior (e.g., Fehr et al., 2002; Gintis, 2011; Gintis et al., 2003; Gintis et al., 2008). But if we juxtapose this method of reasoning with other behavioral experiments, a problem becomes clear. In most experiments, analyses move from real-world observations to isolated motives in experiments and, with newfound data, back to real-world behaviors:

(M1) Observe behavioral pattern in the real world → Discover motives in economic experiments → Make sense of real-world behavior

Strong reciprocity is therefore an unusual mode of scientific inquiry, because it begins with experimental observations rather than real-world behavioral patterns but proceeds to speculate about possible real-world behaviors:

(M2) Observe behavioral pattern in economic experiment → Discover motives in experiment → Make sense of behavior in experiment

What can be concluded definitively from this mode of reasoning is only that people have strongly reciprocal motives in economic games.

Of course, identifying such motives could underscore a real-world behavior or it might just reveal an artifact of economic games. Alternatively, we advocate a middle ground. While experimental evidence for strong reciprocity clearly identifies an impulse for justice in humans, the impulse does not get expressed as strongly in the real world as it does in economic games. A way forward, then, is to investigate the cultural mechanisms that promote or inhibit the impulse in real-world settings.

## Wartime Altruism

One real-world setting in which strong reciprocity is said to be identifiable is among disrupted communities, such as those affected by a natural disaster or warfare, where altruists sustain group cooperation (e.g., Gintis et al., 2008; Mathew & Boyd, 2011). Granting this observation, we provide a few ethnographic examples of wartime altruism, considering whether they are tantamount to strong reciprocity. Wartime altruism is of course a prosocial behavior, and like any other prosocial behavior it involves specific temporal discounting, that is, a regressed time horizon in which greater value is given to the “now” (see Doyle, 2013). In times of war, human beings discount time, often acting in ways that bring immediate reward, for better or worse. Persons in war are known to undertake incredibly unjust and immoral actions against conspecifics, but war also brings out incredible acts of justice and altruism among some individuals. The question is: do such prosocial actions in war constitute strong reciprocity?

### *Costly Punishment in War*

Speaking directly to the role of punishment in promoting cooperation, Mathew and Boyd (2011) examined third-party punishment among the Turkana, an egalitarian, nomadic pastoral society in East Africa. As a group engaging in wartime raids, the Turkana faced significant risks whenever warriors deserted a raiding party. To discourage desertions, the Turkana imposed community-wide sanctions in the form of corporal punishments and fines, which is altruistic since it is a cost that individuals across the group are willing to accept in order to secure justice and cooperation. By taking on such costs, individuals paid a high price alongside others in their

community to ensure that freeriders were punished. Based on a sample of 88 raids, Mathew and Boyd (2011) found that collective third-party punishment significantly lowered desertions and contributed to higher levels of cooperation. As they argue, this example shows that altruistic punishment is significant for small-scale societies but also that negative strong reciprocity could have evolved at the group level in traditional human societies. Granted, the Turkana punish cheaters as a group, which offsets the costs of punishment; though it is done as a group, which differs from economic games, it appears to be compelled by the same moral emotions and with the same consequence of group-level benefits.

Notwithstanding these results, it is difficult to say with certainty that wartime punishment for the Turkana is negative strong reciprocity. After all, even though it is costly, Turkana punishment in war is actually a form of community-wide third-party punishment, which can be more easily rationalized as a costly signal of trustworthiness among community members (Jordan, Hoffman, Bloom, & Rand, 2016). Furthermore, it seems to be an exception to an emerging pattern: when ethnographers of traditional societies (bands, tribes, chiefdoms) detect costly punishment, it is usually second-party punishment, where one is cheated and thereafter avenges oneself; and when third-party punishment does occur, it tends to be collective, thus distributing the costs and risks of doing so (see Lee, 2013/1984, p. 118). Where non-collective third-party punishment is most evident is among persons in state-level societies—but, again, in cross-cultural experiments (e.g., Marlowe et al., 2008). Hence, the Turkana case once again exemplifies the problem of type distinction, and as one of the allegedly best instances of negative strong reciprocity outside of economic games, it is unconvincing.

### *Costly Cooperation in War*

Examples of costly cooperation are more frequent in times of war (e.g., Gintis, 2000a), and they suggest the importance of direct or indirect group-level benefits when communities are disrupted by collective violence. To consider whether these constitute strong reciprocity, we draw from two separate sets of interview data of survivors and ex-fighters of the Yugoslav Wars. The first comes from post-conflict interviews collected by political activist and physician Svetlana Broz (2002), while the second comes from semi-structured interviews collected during 18 months of fieldwork (2015–2016) in the Balkans by Jordan Kiper. What these interviews suggest is that altruistic impulses for what seems to be strong reciprocity are remarkably common in war, as observed by defenders of strong reciprocity (e.g., Gächter & Herrmann, 2009). However, when acted upon, these instances of altruism either fit the descriptions of other evolutionary cooperative behaviors or do not present clear benefits to the reciprocator's group.

When the Yugoslav Wars ended in Bosnia, Broz (2002) began compiling wartime narratives ( $n = 90$ ), with the intent of recording a political history of the wars as told by survivors and ex-fighters (xv–xvi). Besides recording accounts of

war crimes, Broz was surprised to find that many interviewees reported being helped by altruists during the war, often by family, friends, or neighbors—but in some cases by strangers. When Kiper conducted similar interviews with survivors and ex-fighters of the Yugoslav Wars in Croatia, Serbia, and Bosnia Herzegovina ( $n = 174$ ), he was also surprised by the frequency in which interviewees reported being helped by an altruistic stranger. Combining both sets of interviews ( $n = 264$ ), 31 testimonies were about being in a situation of need and receiving help from an unknown person with whom the recipient could not reciprocate. Of these cases, 17 involved being helped by a member of one's ethnoreligious group, but with each of these cases, the altruist was in the company of others and therefore his or her behavior is more accurately characterized as indirect reciprocity or a costly signal to observers. In the remaining 14 cases, the altruist was a stranger from the "other side" of the conflict and, most importantly, put himself or herself at risk by helping, and therefore acted alone and did so in relative secrecy.

Based on these 14 cases, 6 involved a fighter from the other side. These included a fighter protecting someone from being beaten, tortured, or killed ( $n = 2$ ) and helping someone escape from an occupied territory or warzone ( $n = 4$ ). Of the eight cases where a noncombatant helped, interviewees reported being refugees at the time and receiving resources as they fled ( $n = 5$ ), being given rides to escape warzones or pass through enemy checkpoints ( $n = 2$ ), and being hidden from combatants ( $n = 1$ ). We can only speculate as to why persons undertook such costs to help someone who would have been considered their enemy at the time. Perhaps they recognized a family member in the person of need (Broz, 2002), could not stand to see an injustice (p. 371), or simply felt it was the right thing to do (Kiper, unpublished interview data).

Still, the critical question is how this behavior benefits the strong reciprocator's group. One could argue that instead of benefiting their group directly, persons who help outsiders, especially in war, convey the humanity of their own group. Doing so could turn an enemy and thus potential combatant into a sympathetic noncombatant. This sentiment is summarized well by a former Chetnik who was left for dead by his fellow Serbian soldiers after a battle, and then discovered by a Muslim. To the man's surprise, the Muslim did not kill him but rather treated his wounds and took him to a nearby hospital, which, perhaps because the Muslim vouched for the man, accepted him without any questions. Because of the war, the man never found his benefactor and went back to his home once he had healed—but this time as a pacifist. As he reported: "After all I've experience I know there is no force on this earth and no idea that could force me to pick up a gun again" (Broz, 2002, p. 333). Despite this possibility, costly cooperation between would-be enemies in war does not appear to be strong reciprocity. Instead, it is simply another form of general reciprocity, since the group identities of involved parties are known, and the recipient essentially reciprocates with the altruist by forsaking violence against the latter's group.



## ***The Impossibility of Detection***

Our brief analysis of wartime altruism is meant to shed light on what we take to be the fundamental problem of strong reciprocity: even in cases where one would expect to find it, strong reciprocity is difficult to detect with any certainty. As we discussed earlier, this is a problem of type distinction. Any alleged ethnographic instance of strongly reciprocal behavior will blur the lines with other forms of evolved cooperation, which can usually explain the behavior in question with greater clarity and simplicity than strong reciprocity theory. Once again, this problem is rooted in the mode by which strong reciprocity was discovered, that is, as an anomalous behavior within economic games and thereafter sought in the real world, instead of the reverse, which tends to be the common route of investigating a behavior. Likewise, detecting strong reciprocity where we would expect one-shot encounters, such as war, famine, or any other natural disaster, involves real-world problems that often complicate traditional economic theories. For instance, classical economic models assume that humans discount time in a rather consistent way. However, wartime altruism shows that time discounting varies for humans in real-world settings. Detecting the extent of temporal discounting is nevertheless difficult in contexts of war, as people may see their temporal horizon differently, even from moment to moment, depending on their circumstances. Taken together, defenders of strong reciprocity may have to face up to the problem that because the real world cannot match the experimental conditions in which strong reciprocity was discovered, the behavior may be impossible to detect with certainty outside of experimental settings.

## **Final Thoughts**

Our brief discussion of wartime altruism is not intended to assert that strong reciprocity does not exist. Experimental evidence on strong reciprocity suggests that humans indeed have a remarkable inclination for fairness, while cultural group selection provides a sufficient means by which such an inclination would have been selected. Granted that successfully repeated experiments isolate real phenomena and produce materially realized effects (Radder, 2003), experiments on strong reciprocity isolate something real and consequential. What remains partially unanswered, we argue, is the exact nature of strong reciprocity as a phenomenon isolated in experiments, and how that phenomenon changes from the contexts of economic games to the real world. It may no longer be warranted to assume that strong reciprocity in experiments gets expressed as such in the real world, given the lack of concrete ethnographic examples thereof.

We suggest, then, that a narrow interpretation of strong reciprocity may be the best way to move forward. That is to say, researchers should no longer presume that experiments reveal a behavior that one can expect to find in the real world but rather

they isolate a basic psychological or emotional impulse. This impulse underlies the basic human proclivity for fairness and thus justice, which centers on others following or violating social norms, and was probably selected at the group level, just as theorists of strong reciprocity claim. However, much like other naturally selected psychological impulses, the underpinnings of strong reciprocity must be shaped by culture. Consequentially, a potentially rewarding direction for future research is to examine the phenomenology of strong reciprocity and investigate how cultures suppress, cultivate, and manipulate strong reciprocity as a psychological or emotional proclivity to achieve justice. The experimental settings in which strong reciprocity has emerged do not appear to capture the constraints of human social organization, despite the enormous diversity in which humans structure their societies. Strong reciprocity research, therefore, that takes considerations of cultural influences seriously offers a promising approach for understanding the evolution of strong reciprocity and its role in facilitating justice and fairness.

## References

- Alexander, R. (1987). *The biology of moral systems*. New York, NY: Aldine de Gruyter.
- Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal*, 100(401), 464–477.
- Axelrod, R. (1984). *The evolution of cooperation*. New York, NY: Basic Books.
- Axelrod, R., & Hamilton, W. D. (1984). The evolution of cooperation. *Science*, 211, 1390–1396.
- Baldassarri, D. (2013). Prosocial behavior: Evidence from lab-in-the-field experiments. *PLoS One*, 8(3), e58750. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3608652/>.
- Bergmuller, R., & Taboorsky, M. (2010). Animal personality due to social niche specialization. *Trends in Ecology and Evolution*, 25(9), 504–511.
- Boehm, C. (2012). *Moral Origins: The Evolution of Virtue, Altruism, and Shame*. New York: Basic Books.
- Bowles, S., Boyd, R., Matthew, S., & Richerson, P. J. (2012). The punishment that sustains cooperation is often coordinated and costly. *Behavioral and Brain Sciences*, 35(1), 20–21.
- Bowles, S., & Gintis, H. (2002). Homo reciprocans. *Nature*, 415, 125–128.
- Bowles, S., & Gintis, H. (2004). The evolution of strong reciprocity: Cooperation in heterogeneous populations. *Theoretical Population Biology*, 65(1), 17–28.
- Boyd, R., Gintis, H., & Bowles, S. (2010). Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science*, 328(5978), 617–620.
- Boyd, R., Gintis, H., Bowles, S., & Richerson, P. (2003). The evolution of altruistic punishment. *PNAS*, 100(6), 3531–3535.
- Boyd, R., & Richerson, P. J. (2009). Culture and the evolution of human cooperation. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364(1533), 3281–3288.
- Burnham, T., & Johnson, D. (2005). The biological and evolutionary logic of human cooperation. *Analyse & Kritik*, 27, 113–135.
- Brauer, J., & Hanus, D. (2012). Fairness in non-human primates? *Social Justice Research*, 25(3), 256. [http://scholarworks.gsu.edu/cgi/viewcontent.cgi?article=1046&context=psych\\_facpub](http://scholarworks.gsu.edu/cgi/viewcontent.cgi?article=1046&context=psych_facpub).
- Broz, S. (2002). *Good people in an evil time: Portraits of complicity and resistance in the Bosnian war*. New York, NY: Other Press.
- Cialdini, R. (2006). *Influence: The Psychology of Persuasion*. New York, NY: Harper Business.
- Cochran, G., & Harpending, H. (2009). *The 10,000 year explosion: How civilization accelerated human evolution*. New York, NY: Basic Books.

- Cushing, P. J. (1998). Competing the cycle of transformation: Lessons from the rites of passage model. *Pathways: The Ontario Journal of Experimental Education*, 9(5), 7–12.
- Dawkins, R. (2006). *The selfish gene*. Oxford: Oxford University Press. (Original work published in 1976).
- de Quervain, D. J., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., & Buck, A. (2004). The neural basis of altruistic punishment. *Science*, 305, 1254–1258.
- Debove, S., Baumard, N., & Andre, J. B. (2016). On the evolutionary origins of equity. *BioRxiv*. <http://dx.doi.org/10.1101/052290>.
- Diekmann, A., Jann, B., Przepiorka, W., & Wehrl, S. (2014). Reputation and the evolution of cooperation in anonymous online markets. *American Sociological Review*, 79(1), 65–85.
- Doyle, J. (2013). Survey of time preference, delay discounting models. *Judgment and Decision making*, 8(2), 116–135.
- Dugatkin, L. A. (1999). *Cheating monkeys and citizen bees*. New York, NY: Simon & Shuster.
- Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, 14(4), 583–610.
- Fehr, E., Fischbacher, U., & Gächter, S. (2002). Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature*, 13(1), 1–25.
- Fehr, E., & Gächter, S. (2001). Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4), 980.
- Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economic Letters*, 71(3), 397–404.
- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415, 6868.
- Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, 425, 785–791.
- Fehr, E., & Rockenbach, B. (2003). Detrimental effects of sanctions on human altruism. *Nature*, 422, 137–140.
- Fehr, E., & Henrich, J. (2003). Is strong reciprocity a maladaptation? On the evolutionary foundations of human altruism. In P. Hammerstein (Ed.), *Genetic and Cultural Evolution of Cooperation* (pp. 55–82). Cambridge, MA: MIT Press.
- Fehr, E., & Fischbacher, U. (2004). The nature of human altruism. *Nature*, 425, 785–791.
- Fehr, E., & Leibbrandt, A. (2011). A field study on cooperativeness and impatience in the tragedy of the commons. *Journal of Public Economics*, 95(10), 1144.
- Feldman, R. (2015). Mutual influences between child emotion regulation and parent-child reciprocity support development across the first 10 years of life: Implications for developmental psychopathology. *Development and Psychopathology*, 27(1), 1007–1023.
- Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, 71(3), 397–404.
- Gächter, S., & Herrmann, B. (2009). Reciprocity, culture, and human cooperation: Previous insights and a new cross-cultural experiment. *Philosophical Transactions of the Royal Society Biological Sciences*, 364, 791–806.
- Gintis, H. (2000a). Strong reciprocity and human sociality. *Journal of Theoretical Biology*, 206, 169–179.
- Gintis, H. (2000b). Group selection and human prosociality. *Journal of Consciousness Studies*, 7(1), 215–219.
- Gintis, H. (2003). The hitchhiker's guide to altruism: Genes, culture, and the internalization of norms. *Journal of Theoretical Biology*, 220(4), 407–418.
- Gintis, H. (2011). Gene-culture coevolution and the nature of human sociality. *Philosophical Transactions of the Royal Society Biological Sciences*, 366, 878–888.
- Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (2003). Explaining altruistic behavior in humans. *Evolution and Human Behavior*, 24, 153–172.
- Gintis, H., & Fehr, E. (2012). The social structure of cooperation and punishment. *Behavioral and Brain Sciences*, 35(1), 28–29.
- Gintis, H., Henrich, J., Bowles, S., Boyd, R., & Fehr, E. (2008). Strong reciprocity and the roots of morality. *Social Justice Research*, 21(2), 241–253.
- Grafen, A. (1985). A geometric view of relatedness. *Oxford Survey of Evolutionary Biology*, 2, 28–89.

- Grusec, J. E., & Kuczynski, L. (1997). *Parenting and Children's internationalization of values: A handbook of contemporary theory*. New York, NY: John Wiley & Sons.
- Guala, F. (2012). Reciprocity: Weak or strong? What punishment experiments do (and do not) demonstrate. *Behavioral and Brain Sciences*, 35(1), 1–59.
- Guiso, L., Sapienza, P., & Zingales, L. (2009). *The Quarterly Journal of Economics*, 124(3), 1095–1131.
- Hagen, E., & Hammerstein, P. (2006). Game theory and human evolution: A critique of some recent interpretations of experimental games. *Theoretical Population Biology*, 69, 339–348.
- Hamilton, W. D. (1964). The genetical evolution of social behavior I & II. *Journal of Theoretical Biology*, 7(1), 1–52.
- Henrich, J., & Boyd, R. (2001). Why people punish defectors: Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *Journal of Theoretical Biology*, 208, 79–89.
- Henrich, J., & Chudek, M. (2012). Understanding the research program. *Behavioral and Brain Sciences*, 35(1), 29–30.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., & Gintis, H. (2004). *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*. New York: Oxford University Press.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33, 61–83.
- Henrich, J., Boyd, R., Camerer, C., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., Ensminger, J., Henrich, N.S., Hill, K., Gil-White, F., Gurven, M., Marlowe, F.W., Patton, J.Q., & Tracer, D. (2005). 'Economic man' in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences*, 28(6), 795–815.
- Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., ... Zilker, J. (2006). Costly punishment across human societies. *Science*, 312, 1767–1770.
- Hsu, M., Anen, C., & Quartz, S. R. (2008). The right and the good: Distributive justice and neural encoding of equity and efficiency. *Science*, 320, 1092–1095.
- Inglis, F., West, S., & Buckling, A. (2014). An experimental study of strong reciprocity in bacteria. *Biological Letters*, 10, 20131069.
- Jordan, J., Hoffman, M., Bloom, P., & Rand, D. (2016). Third-party punishment as a costly signal of trustworthiness. *Nature*, 530, 473–476.
- Kaltwasser, L., Hildebrandt, A., Wilhelm, O., & Sommer, W. (2016). Behavioral and neuronal determinants of negative reciprocity in the ultimatum game. *Social Cognition and Affective Neuroscience*, 11(11), 1608–1617.
- Knoch, D., Nitsche, M. A., Fischbacher, U., Eisenegger, C., Pascual-Leone, A., & Fehr, E. (2008). Studying the neurobiology of social interaction with transcranial direct current stimulation—The example of punishing unfairness. *Cerebral Cortex*, 18, 1987–1990.
- Li, Y., & Yamagishi, T. (2014). A test of the strong reciprocity model: A relationship between cooperation and punishment. *Shinrigaku Kenkyu*, 85(1), 100–105.
- Lee, R. (2013). *The Dobe Ju/'hoansi*. Belmont, CA: Wadsworth. (Original work published 1984).
- Leland, K. N., & Brown, G. R. (2002). *Sense and nonsense: Evolutionary perspectives on human behavior*. Oxford: Oxford University Press.
- Luo, Q., Nakic, M., Wheatley, T., Ritchell, R., Martin, A., & Blair, R. J. (2006). The neural basis of implicit moral attitude—An IAT study using event-related fMRI. *NeuroImage*, 30, 1449–1457.
- Marlowe, F. W., Berbesque, C., Barr, A., Barrett, C., Bolyanatz, A., Camilo, J., ... Tracer, D. (2008). More 'altruistic' punishment in larger societies. *Proceedings of the Royal Society Biological Sciences*, 275, 587–592.
- Mathew, S., & Boyd, R. (2011). Punishment sustains large-scale cooperation in prestate warfare. *Proceedings of the National Academy of Sciences of the United States of America*, 108(28), 11375–11380.
- Mauss, M. (1990). *The gift: The form and reason for exchange in archaic societies*. New York, NY: W.W. Norton & Company. (Original work published in 1950).

- Maynard Smith, J. (1964). Group selection and kin selection. *Nature*, 201, 1144–1147.
- Mendez, M. F. (2009). The neurobiology of moral behavior: Review and neuropsychiatric implications. *CNS Spectrums*, 14(11), 608–620.
- Moll, J., Zahn, R., de Oliveira-Souza, R., Krueger, F., & Grafman, J. (2005). The neural basis of human moral cognition. *Nature Reviews Neuroscience*, 6, 799–809.
- Moll, H., & Tomasello, M. (2007). Cooperation and human cognition: The Vygotskian intelligence hypothesis. *Philosophical Transactions of the Royal Society Biological Sciences*, 362(1480), 639–648.
- Nisbett, R. E., & Cohen, D. (1996). *Culture of honor: The psychology of violence in the south*. Boulder, CO: Westview Press.
- Nowak, M., & Sigmund, K. (1998). Evolution of indirect reciprocity. *Nature*, 393, 573–577.
- Nowak, M., Tarnita, C. E., & Wilson, E. O. (2010). The evolution of eusociality. *Nature*, 466(7310), 1057–1062.
- Pascual, L., Rodrigues, P., & Gallardo-Pujol, D. (2013). How does morality work in the brain? A functional structural perspective of moral behavior. *Frontiers in Integrative Neuroscience*, 7(65), 1–8.
- Pisor, A. C., & Fessler, D. M. (2012). Importing social preferences across contexts and the pitfall of over-generalization across theories. *Behavioral and Brain Sciences*, 35(1), 34–35.
- Price, M. E. (2008). The resurrection of group selection as a theory of human cooperation. *Social Justice Research*, 21, 228–240.
- Radder, H. (2003). *The philosophy of scientific experimentation*. Pittsburgh, PA: University of Pittsburgh Press.
- Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University Press.
- Rustagi, D., Engel, S., & Kosfeld, M. (2010). Conditional cooperation and costly monitoring explain success in forest commons management. *Science*, 330(6006), 961–965.
- Sahlins, M. (1972). *Stone Age Economics*. Chicago, IL: Aldine-Atherton.
- Sanfey, A. G., Rilling, J. K., Aaronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, 300, 1755–1758.
- Schneider, F., & Fehr, E. (2010). Eyes are watching but nobody cares: The irrelevance of eye cues for strong reciprocity. *Proceedings of the Royal Society of London: B-Biological Sciences*, 277, 1315–1323.
- Sober, E., & Wilson, D. S. (1998). *Unto others: The evolution and psychology of unselfish behavior*. Cambridge, MA: Harvard University Press.
- Sosis, R. (2006). Religious behaviors, badges, and bans: Signaling theory and the evolution of religion. In P. McNamara (Ed.), *Where god and science meet: How brain and evolutionary studies alter our understanding of religion* (pp. 61–68). Westport, CT: Praeger.
- Sterelny, K. (2011). From hominins to humans: How *sapiens* became behaviourally modern. *Philosophical Transaction of the Royal Society: B-Biological Sciences*, 366(1566), 809–822.
- Tabibnia, G., Satpute, A. B., & Lieberman, M. D. (2008). The sunny side of fairness: Preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychological Science*, 19, 339–347.
- Tracer, D. (2003). Selfishness and fairness in economic and evolutionary perspective: An experimental economic study in Papua New Guinea. *Current Anthropology*, 44(3), 432–443.
- Trivers, R. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, 46(1), 35–57.
- Trivers, R. (2006). Reciprocal altruism: 30 years later. In P. M. Kappeler & C. P. van Shaik (Eds.), *Cooperation in primates and humans* (pp. 67–84). New York, NY: Springer-Verlag Berlin.
- Watanabe, T., Takezawa, M., Nakawake, Y., Kunimatsu, A., Yamasure, H., Nakamura, M., ... Masuda, N. (2014). Two distinct neural mechanisms underlying indirect reciprocity. *Proceedings of the National Academy of Sciences of the United States of America*, 111(11), 3990–3995.
- West, S. A., Mouden, C. E., & Gardner, A. (2011). Sixteen misconceptions about the evolution of cooperation in humans. *Evolution and Human Behavior*, 32, 231–262.

- Williams, G. (1966). *Adaptation and Natural Selection*. Princeton, NJ: Princeton University Press.
- Wilson, D. S., & Sober, E. (1994). Reintroducing group selection to the human behavioral sciences. *Behavioral and Brain Sciences*, *17*(4), 585–654.
- Wynne-Edwards, V. C. (1962). *Animal dispersion in relation to social behavior*. Edinburgh: Oliver & Boyd.
- Wynne-Edwards, V. C. (1964). Group selection and kin selection: Reply to Maynard Smith. *Nature*, *201*, 1147.
- Yamagishi, T., Horita, Y., Mifune, N., Hashimoto, H., Li, Y., Shinada, M., ... Simunovic, D. (2012). Rejection of unfair offers in the ultimatum game is no evidence of strong reciprocity. *Proceedings of the National academy of Sciences of the United States of America*, *109*(50), 20364.
- Zahavi, A. (1975). Mate selection—A selection for a handicap. *Journal of Theoretical Biology*, *53*, 205–214.