The Reciprocity Controversy

Gerald Carter

1University of Maryland

*Corresponding author (Email: gcarter@umd.edu)


Abstract - Reciprocity (or “reciprocal altruism”) was once considered an important and widespread evolutionary explanation for cooperation, yet many reviews now conclude that it is rare or absent outside of humans. Here, I show that nonhuman reciprocity seems rare mainly because its meaning has changed over time. The original broad concept of reciprocity is well supported by evidence, but subsequent divergent uses of the term have relied on various translations of the strategy ‘tit-for-tat’ in the repeated Prisoner’s Dilemma game. This model has resulted in four problematic approaches to defining and testing reciprocity. Authors that deny evidence of nonhuman reciprocity tend to (1) assume that it requires sophisticated cognition, (2) focus exclusively on short-term contingency with a single partner, (3) require paradoxical evidence for a temporary lifetime fitness cost, and (4) assume that responses to investments are fixed. While these restrictions basically define reciprocity out of existence, evidence shows that fungi, plants, fish, birds, rats, and primates enforce mutual benefit by contingently altering their cooperative investments based on the cooperative returns, just as predicted by the original reciprocity theory.

Keywords - Cooperation, Prisoner’s Dilemma, Pseudoreciprocity, Reciprocity, Reciprocal altruism

Comparative psychologists, evolutionary psychologists, and behavioral ecologists often study cooperation using different theories and methods, asking questions at different levels of analysis. What cues trigger the cooperative behavior? How does it develop? When did it evolve? Why is it adaptive? The multidisciplinary nature of this field leads to new connections but also miscommunication. For instance, some semantic confusion occurs because comparative psychologists often define behaviors such as ‘cooperation’, ‘altruism’, and ‘mutualism’ based on proximate goals or motivations, similar to their colloquial usage (de Waal, 2008), whereas evolutionary biologists define these terms based on the ultimate net effects on direct fitness (i.e. lifetime reproductive success, West, Griffin, & Gardner, 2007a,b). Many misunderstandings resulting from these semantic discrepancies have been resolved elsewhere (see Noë, 2006, West, El Mouden, & Gardner, 2011; West et al., 2007b), but one important concept that continues to cause confusion is ‘reciprocity’ (or ‘reciprocal altruism’ Trivers, 1971).

Reciprocity is one of the best-known evolutionary explanations for cooperation, but also among the most controversial (Cheney, 2011; Clutton-Brock, 2009; Hammerstein, 2003; Schino & Aureli 2010a,b). Although once considered the key explanation for helping between non-kin, most reviews now conclude that it is absent or very rare outside of humans (e.g., Clutton-Brock, 2009; West et al., 2011). All claims of reciprocity have been disputed, including experimental evidence from fish (reviewed by Dugatkin, 1997), rodents (Rutte & Taborsky, 2008), birds (Krama et al., 2012; Krams, Krama, Igaune, & Mänd, 2008; Krams et al., 2013), and primates (reviewed by de Waal & Suchak, 2010; Schino & Aureli 2008, 2009). As a consequence, theorists attempt to explain why reciprocity is so rare (André, 2014), while others view reciprocity as an important and underappreciated mechanism for cooperation (Schino & Aureli, 2010a,b; Taborsky, 2013).
Reciprocity assumes that cooperative investments can be exploited if the recipients do not provide adequate cooperative returns back to the actor (i.e., ‘cheating’ Ghoul, Griffin, & West, 2013), and it predicts that individuals will therefore adjust these investments contingent on the returns received from their partners. Some authors contrast ‘direct reciprocity’ (A helps B because B helps A) with ‘indirect reciprocity’ (A helps B because B helps C) or ‘generalized reciprocity’ (A helps B because A was helped), and some authors separate positive reciprocity (contingent reward) from negative reciprocity (contingent punishment). The tendency of humans to both cooperate and punish non-cooperators, even at a cost or in one-shot economic games, has been called ‘strong reciprocity’ (reviewed by West et al., 2007b, 2011). Here, I focus exclusively on ‘direct reciprocity’ and do not distinguish between positive and negative effects. Relevant terms are defined in Box 1.

**Box 1. Glossary**

**Altruism:** cooperation that on average decreases the actor’s direct fitness.

**Byproduct mutualism:** mutual benefits that are incidental (the traits or behaviors were not shaped by natural selection to provide benefits to others).

**Cooperation:** a behavior or trait that on average increases the inclusive fitness of both the actor and the recipient; includes altruism and mutualism.

**Cooperative investment:** an action that aids a recipient and functions to provide a cooperative return to the actor.

**Cooperative return:** an action by a recipient of a cooperative investment that increases the investor’s direct fitness.

**Direct fitness:** lifetime reproductive success; number of total offspring that survive until adulthood.

**Enforcement mechanism:** a behavior or ability that functions to ensure that cooperative investments yield an indirect or direct fitness return (enforcement prevents cheating).

**Cheating:** occurs when a cooperative investment decreases the helper’s inclusive fitness (the recipients do not provide a cooperative return or are not the intended recipients).

**Inclusive fitness:** the sum of direct and indirect fitness (traits are adaptive when they increase inclusive fitness).

**Direct fitness:** the component of inclusive fitness gained from helping relatives.

**Mutualism:** cooperation that on average increases the direct fitness of the actor and recipient.

**Pseudoreciprocity:** unconditional cooperative investments that enable an inevitable byproduct return (no cheating and no enforcement).

**Reciprocity:** contingent cooperative investments that are based on the cooperative returns (enforcement through partner control and/or partner choice).

The reciprocity controversy depends more on semantic disagreements than on disputes about observable behavior or social evolution theory. Whereas the original concept of reciprocity was broad (Axelrod & Hamilton, 1981; Trivers, 1971), operational definitions gradually diverged and became narrower in meaning, but these multiple definitions are now used interchangeably, resulting in confusion. Most studies of reciprocity have tested if the short-term payoffs of a given scenario in nature fit the Prisoner’s Dilemma game structure and if the behavior of organisms follows the strategy “tit-for-tat” (cooperate, then copy partner’s last move). This model of reciprocity has led to four problematic distinctions that have essentially defined ‘reciprocity’ out of existence.

First, some animal behavior researchers have taken the play of economic games very literally and assume that reciprocity is an intentional strategy requiring an understanding of game payoffs and the ability to keep score, plan ahead, and delay gratification (I call this ‘the calculated reciprocity error’). Second, some operational definitions focus exclusively on short-term contingency with a single partner while ignoring factors such as partner choice, power asymmetries, and foundations of prior experience (‘the short-term contingency bias’). Third, some definitions require demonstrating that an adaptive helping behavior reduces lifetime fitness but only in the short-term (‘the temporary fitness cost paradox’). Finally, endless controversy concerns whether the returns on a cooperative investment are costly and strategically enforced (reciprocity) or self-serving and inevitable (pseudoreciprocity), a distinction that can be semantic (‘the byproduct ambiguity’). To understand how these issues have arisen, we must take a historical perspective.

**Historical Background**
Robert Trivers (1971) coined the term ‘reciprocal altruism’ to explain how apparently altruistic behavior could evolve between non-kin. ‘Reciprocal altruism’ is not a form of altruism in the evolutionary sense (sensu Hamilton, 1964) because it does not decrease lifetime direct fitness. Many authors therefore prefer the term ‘reciprocity’ (Alexander, 1974; Axelrod & Hamilton, 1981; West et al., 2007a,b). The choice of the term ‘reciprocal altruism’ has likely led to at least some of the confusion that reciprocity is an alternative to mutual benefit, rather than a way of enforcing it (West et al., 2007b).

Trivers (1971) defined ‘reciprocal altruism’ in both narrow and broad terms. In the narrowest sense, he described various ways that it could operate in humans. In the broadest sense, he considered almost any case of a delayed mutual benefit to support the theory. For example, he imagined a hypothetical scenario in which a bird benefits from alarm calling because the act somehow makes a predator less likely to target the caller in the future. In his description, any social benefit to other birds was an incidental byproduct, a form of cooperation now called byproduct mutualism (Brown, 1983; Connor, 1986, 1995a; West-Eberhard, 1975). Modern usage of the term reciprocity excludes simple byproduct mutualisms, and Trivers (2006) later clarified this point himself.

Axelrod and Hamilton (1981) modeled reciprocity using the simple strategy “tit for tat” in the repeated Prisoner’s Dilemma. In this game, each player chooses to ‘cooperate’ or ‘defect’ and receives a different payoff depending on the other’s simultaneous response. Four payoffs are possible: both players cooperate (R, reward for mutual cooperation), cooperate with defector (S, sucker’s payoff), defect against cooperator (T, temptation to defect), or both players defect (P, punishment for mutual defection). In the single-shot Prisoner’s Dilemma, T > R > P > S, and defection is the only stable strategy. If the game is repeated in a series of continual rounds, then “tit for tat” (cooperate on first round, then copy player’s last move) can outcompete “always defect” and many other strategies.

This model led to an explosion of subsequent game theory models for cooperation. Reciprocity was soon equated with both game theory and tit for tat. Yet it was still unclear exactly what constituted empirical evidence for reciprocity and how best to translate game theory to experiment. The ambiguity in defining reciprocity led to a workshop meeting where leading researchers concluded that reciprocity “might be very rare and restricted to a few groups, or it might be quite common and widespread – this depends on how the phenomenon is defined and the importance attributed to animals’ intentions” (Taylor & McGuire, 1988, p. 69). Axelrod and Hamilton (1981) considered reciprocity broadly applicable to cooperation between neighboring male songbirds, interspecific mutualisms, microbes, viruses, and even chromosomes.

Several authors argued that the original definition of reciprocity was too broad (e.g., Koenig, 1988; Waltz 1981). These researchers thought the term had become too inclusive, because behaviors such as monogamy (Ligon 1983), mutual restraint of aggression (Lombardo, 1985), and sex (West-Eberhard, 1975) were being labeled as reciprocal altruism or reciprocity. In response, they argued that ‘reciprocal altruism’ should be used only for acts of helping that pose fitness costs to the helper (Koenig, 1988; Wilkinson, 1988).

By the 1990s, some reviews claimed that reciprocity was common (e.g., Dugatkin, 1997) while others argued that it was rare (e.g., Clements & Stephens, 1995). In most cases, the controversy involved whether a particular behavior actually conforms to tit-for-tat in the Prisoner’s Dilemma (reviewed by Noë, 2006; Pusey & Packer, 1997; Raihani & Bshary, 2011). By the mid-2000s, interest in the Prisoner’s Dilemma game as a model for cooperation had begun to decline due to the difficulty in translating theory to reality (Noë, 2006, Raihani & Bshary, 2011). Trivers (2006) lamented that:

Theorists and empiricists alike were forgetting that iterated games of PD amount to a highly artificial model of social interactions; each successive interaction simultaneous, costs and benefits never varying, options limited to only two moves, no errors, no escalated punishment, no population variability within traits and so on. In fact, almost all of these simplifying assumptions have now been shown to introduce important effects. (p. 70).
As game-theoretical models grew increasingly detached from empirical work (e.g., Nowak, 2006), the term reciprocity, now associated with such models, fell out of favor with behavioral ecologists. As expressed by West et al. (2007a):

We do not need more convoluted theoretical analyses of games such as the Prisoner’s Dilemma, snow drift, etc. … games such as the Prisoner’s Dilemma and its solution with various forms of reciprocity make a large number of extremely specific and often unrealistic assumptions. (p. R669).

Yet at the same time, work on interspecific mutualisms was accumulating a great deal of evidence that cooperative investments are indeed contingent on variable cooperative returns (Box 2). But the term ‘reciprocity’ was rarely used here. Instead, researchers referred to ‘sanctions’ (Kiers, Rosseau, West, & Denison, 2003), ‘reciprocal rewards’ (Kiers et al., 2011), and ‘partner choice’ (Noë & Hammerstein, 2001). Reciprocity is now largely equated with intraspecific, rather than interspecific cooperation even though it was applied originally to both (Axelrod & Hamilton, 1981; Trivers, 1971).

Experimental studies on cooperative exchanges among fish, plants, fungi and bacteria have tested the behavioral response to simulated cheating by making one partner able to receive, but not reciprocate, a cooperative investment (e.g., Kiers et al., 2003, 2011). The results of such studies have shown that partner choice, partner switching, and partner control (reward and punishment of a single partner), as well as various byproduct benefits that depend on ecological circumstances, can all play key roles in stabilizing cooperation (Box 2). These results clearly illustrate that enforcement mechanisms are often necessary to stabilize cooperation and complex cognition is not required for sophisticated mechanisms of partner control or choice.

Such studies also demonstrated the utility of viewing cooperation using the metaphor of investment, exchange, supply, and demand. This approach was developed by biological market theory (Noë & Hammerstein, 1994, 1995, 2001) and has provided some of the clearest predictions regarding cooperation both between and within species (e.g., Fruteau, Voelkl, Van Damme, & Noë, 2009; Kiers et al., 2011). Biological market models have now largely replaced the prisoner’s dilemma and other game theory models for guiding empirical studies of what used to be described as reciprocity.

**Box 2. Examples of contingent cooperative exchanges in intraspecific mutualisms**

In the cleaner-client fish mutualism, small cleaners cooperatively eat dead skin off larger ‘client’ fish, but can also ‘cheat’ by eating mucus or live tissue (Grutter, 1999). Both cleaners and clients enforce cooperation. Clients abandon or punish cleaners that cheat and avoid cleaners that they observe cheating (Bshary & Grutter, 2002a, 2005, 2006). Cleaners remember the time, location, and quality of client interactions (Salwiczek & Bshary, 2011), behave more cooperatively when observed by non-resident clients (Bshary & Grutter, 2006, Pinto, Oates, Grutter, & Bshary, 2011), increase cooperation by punishing other cleaners (Bshary, Grutter, Willemer, & Leimar, 2008, Raihani, Grutter, & Bshary, 2010), and adjust the extent of third-party punishment to client value and the size of conspecific partners (Raihani, Pinto, Grutter, Wismer, & Bshary, 2012).

Plants exchange resources with several symbiotic partners, including mycorrhizal fungi, rhizobia bacteria, and pollinating insects. By diverting resources to different structures, plants selectively kill symbionts that do not provide returns (‘sanctions’, Kiers et al., 2003; see also Goto, Okamoto, Toby Kiers, Kawakita, & Kato, 2010; Jandér & Herre, 2010). In other cases, contingent enforcement is reciprocal; in the plant-mycorrhizal fungi mutualism, both partners reward high returns and punish low returns (Hammer, Pallon, Wallander, & Olsson, 2011; Kiers et al., 2011). Importantly, the contingent investments are often continuous rather than discrete such that the intensity of sanctions matches the amount of the return (Kiers, Rosseau, & Denison, 2006).

In the acacia-ant mutualism, a host plant exchanges nectar for defense by ants. Even before enforcement is considered, cheating is already inhibited by a byproduct benefit; the aggressiveness of ants is linked to both their ability to defend host plants and to outcompete less aggressive ant species (Heil, 2013). However, plants still possess several enforcement strategies. They produce nectars that are difficult to digest for non-mutualists (Orona-Tamayo et al., 2013), and these nectars also manipulate the digestive system of their ant mutualists towards dependency on the nectar rewards (Heil, Barajas-Barron, Orona-Tamayo, WielSch, & Svatos, 2103). On the other end, ant strategies of partner control appear to include contingent defense of plants based on amount of nectar supplied (Orona-Tamayo & Heil, 2013).
The Calculated Reciprocity Error

Recent authors have argued that reciprocity requires sophisticated cognitive abilities for making planned intentional investments with an expectation of a future reward (e.g., Hauser, McAuliffe, & Blake, 2009; Ramseyer, Pelé, Dufour, Chauvin, & Thierry, 2006; Russell & Wright, 2009; Stevens, Cushman, & Hauser, 2005; Stevens & Hauser, 2004). For example, Stevens and Hauser (2004) stress that reciprocity is potentially limited by capacities for “numerical discrimination, time estimation, delayed gratification, detection and punishment of cheaters, analysis and recall of reputation, and inhibitory control.” This ‘calculated reciprocity’ (de Waal & Luttrell, 1988) leads to an operational definition that requires testing that an animal can strategically resist the temptation to defect to obtain a delayed social reward, even under extremely artificial conditions. For instance, experiments found that blue jays did not learn to perform a tit for tat strategy in an operant conditioning paradigm that mimicked a Prisoner’s Dilemma in the absence of any natural or social cues (Clements & Stephens, 1995; Stephens, McLinn, & Stevens, 2002, reviewed by Noë, 2006). Evidence for calculated reciprocity in nonhuman animals under these conditions is rare (e.g., Hauser et al., 2009, but see Dufour, Pelé, Neumann, Thierry, & Call, 2009). This evidence has been used to suggest that reciprocity might be rare in nature, but this conclusion assumes that all reciprocity is calculated reciprocity and acquired through associative learning.

The alternative view is that the ‘calculations’ required for reciprocity occur not via associative learning alone, but through task-specific adaptations, which require the proper ecologically relevant cues to act as triggers. According to this view, reciprocity occurs as a species-specific cognitive specialization similar to evolved mechanisms for mate choice, navigation, or kin recognition. For example, the extraordinary species-specific abilities of food-caching birds to remember thousands of cache locations over months is not constrained by the supposed difficulties of long-term memory, delaying gratification, and planning for the future (Bednekoff, Balda, Kamil, & Hile, 1997).

With this in mind, Stevens et al. (2005) acknowledged that “we should expect to find reciprocity and punishment in instances where adaptation has overcome the initial cognitive constraints – where narrowly tailored cognitive mechanisms have evolved to support specific behavioral routines (p. 512).” The controversy over the cognitive constraints on reciprocity therefore rests on deeper controversies over how easily adaptation overcomes cognitive constraints and how often social behaviors rely on context-specific adaptive specializations rather than on associative learning (e.g., Magpahil & Bolhuis, 2001).

An adaptationist view is that associative learning cannot fully explain reciprocity. Consider that kin discrimination (which is often based on prior association) requires different adaptive designs for different taxa. This results not only from physical constraints (e.g., plants don’t have brains), but also differing ecological requirements (e.g., location-based offspring recognition can work for stationary bank swallow nestlings but not mobile penguin chicks). Even when reciprocity is based on learning the relative payoffs of helping through operant conditioning, this learning process will likely be shaped by natural selection, such that the task will be acquired faster in species performing reciprocity. This prediction is consistent with the finding that adult cleaner fish outperform juvenile cleaners and several primates at learning a cooperative task that simulates the payoffs that cleaners regularly face in nature (Salwiczek et al., 2012).

Calculated reciprocity in humans. Even in humans, calculated reciprocity in humans often appears ‘instinctive’, subconscious, and context-specific. Rather than relying on strategic self-control, many human prosocial behaviors are fast, intuitive, and built into our basic emotions (Frank, 1988; Trivers, 1971). Reasoning through a logic puzzle is slow and difficult compared to the way insight is quickly gained about the same logical problem framed as a social exchange (e.g., Cosmides, 1989; Cosmides & Tooby, 1992). Testing calculated reciprocity-- by placing people in Prisoner’s Dilemma or other economic games—often leads to irrational decisions which appear to reflect decisions that would be optimal under more natural circumstances (Burton-Chellew & West, 2012, 2013). Humans treat single-shot economic games as if they might be repeated (Deltow, Krasnow, Cosmides, & Tooby, 2011). Cooperative outcomes in the Prisoner’s Dilemma are inherently more rewarding and activate distinct...
reward regions in the brain when the payoffs occur with a human partner rather than with a computer (Abric & Kahan, 1972; Rilling et al., 2002). In stark contrast to avoiding a temptation to defect, most defectors feel an initial impulse to cooperate (Rand, Greene, & Nowak, 2012). Cooperative decisions to donate to public goods are influenced by irrational audience cues (e.g., pictures of eyes, Bateson, Nettle, & Roberts, 2006; Haley & Fessler, 2005) or cues to group competition (Burton-Chellew & West, 2012). Such findings only make sense if many heuristics for cooperative decision-making subconsciously rely on cues that would have maximized inclusive fitness in ancestral environments. In short, the payoffs as given by an experiment are not always the payoffs that are perceived by animal minds.

Similarly, social birds and mammals probably engage different neurocognitive mechanisms when learning that food can be obtained by grooming others versus pecking keys. The importance of ecological and social cues is therefore extremely relevant for reciprocity tests in both human and animal subjects. This likely explains why reciprocity experiments in nonhuman primates are more likely to find evidence for short-term contingency when the experimenters test natural helping behaviors in a group setting rather than use artificial designs with paired subjects performing instrumental tasks (Jaeggi, De Groot, Stevens, & Van Schaik, 2012).

The Short-term Contingency Bias

There is abundant and growing evidence for symmetrical patterns of helping at the group level (‘symmetry-based reciprocity,’ de Waal & Luttrell, 1988), which are consistent with reciprocity but not kinship biases (e.g., bats: Carter & Wilkinson, 2013a,b,c; corvids: Fraser & Bugnyar, 2012; Scheid et al., 2008; primates: Gomes, Mundy, & Boesch, 2009; Schino & Aureli, 2008). However, such correlations tell us little about causation.

Experimental studies have historically emphasized short-term alternation of helping acts with a single partner, especially in primates (reviewed by de Waal & Brosnan, 2006; de Waal & Suchak, 2010; Schino & Aureli, 2009). For example, ‘attitudinal reciprocity’ (de Waal, 2000), relies on emotional scorekeeping, but is defined as when ‘parties mirror each other’s social attitudes with a high degree of short-term contingency’ (de Waal & Suchak, 2010, Table 1). However, an overemphasis on short-term alternating exchange can ignore the roles of prior long-term social relationships and partner choice.

Contingency in a human friendship. Studies of how reciprocity works in humans can guide our expectations about what to expect in other primates or vertebrates. Whereas calculated reciprocity is used in human trade, most human social relationships (e.g. communal relationships, Clark & Mills, 1979) are likely enforced by attitudinal reciprocity. Trivers (1971) used reciprocity to explain friendship and moral emotions such as guilt, shame, gratitude, sympathy, and trust. But subsequent authors (e.g., Silk, 2003) have suggested that reciprocity cannot explain friendship because friends do not appear to closely track favors (the calculated reciprocity error). Humans express a stronger obligation to repay favors to strangers, while exchanges of goods or services in human friendships are often implicit, delayed, imprecise, and even offensive and taboo if they are explicit (Boster, Rodriguez, Cruz, & Marshall, 1995; Shackelford & Buss, 1996; Silk 2003). Why might this be?

One explanation is that a desire to immediately repay social debt signals that future interactions are not expected. Concealing expectations of ‘exchange’ might also function similarly to indirect speech (Pinker, Nowak, & Lee, 2008): it allows people to negotiate topics of implicit social conflict while maintaining plausible deniability about their own expectations. Put differently, friends do not discuss long-term exchange of social support services for the same reason that dating does not involve explicit discussion of reproduction.

Although often implicit, reciprocity is clearly embedded within the psychology of human friendships; social investments are affected by changes in the ability of friends to reciprocate, the availability of alternative friends, and the need for social support (e.g., Cosmides & Tooby, 1992; Frank, 1988; Shackelford & Buss, 1996; Tooby & Cosmides, 1996). Humans tolerate short-term imbalances with friends more than strangers and track the cooperative acts of strangers more than friends, but they
still track the investments of friends (Xue & Silk, 2012). The same can likely be said for other kinds of social relationships such as between spouses or siblings.

**Contingency in a long-term animal relationship.** Many nonhuman animals possess long-term cooperative social bonds that are functionally analogous to human friendships. Such long-term cooperative social bonds (henceforth “social bonds”) are well described in chimpanzees and baboons (Seyfarth & Cheney, 2012), and there is also evidence for their existence in macaques, capuchin monkeys, elephants, feral horses, hyena, dolphins, bats, corvids, and mice (Braun & Bugnyar, 2012; Carter & Wilkinson, 2013c; Fraser & Bugnyar, 2012; Weidt, Hofmann, König, 2008; Weidt, Lindholm, & König, 2014). Field studies have demonstrated that strong social bonds provide clear fitness benefits (e.g., Cameron, Setsaas, & Linklater, 2009; Schülke, Bhagavatula, Vigilant, & Ostner, 2010; Silk et al., 2010).

Long-term social bonds are often better than recent social experience at predicting cooperative investments (Carter & Wilkinson, 2013a,c; Gomes & Boesch, 2011; Sabbatini, Vizioli, Visalberghi, & Schino, 2012; Seyfarth & Cheney, 2012). Partner fidelity through social bonding reduces social risks and facilitates the exchange of multiple cooperative investments such as information transfer, social thermoregulation and grooming, cooperative foraging and food sharing, and protection from predators and hostile conspecifics. Several primatologists have recently outlined how implicit knowledge of social relationships can simplify the process of reciprocity by reducing these multiple currencies of help into a single trackable currency of relationship quality (de Waal, 2000; Jaeggi et al., 2012; Massen, Sterck, & de Vos 2010; Schino & Aureli, 2009, 2010a,b; Seyfarth & Cheney, 2012; Tooby & Cosmides, 2008). For example, chimpanzees of both sexes appear to exchange several different commodities, including grooming, sex, support, and food, resulting in balanced long-term relationships (Gomes & Boesch, 2011). As Seyfarth & Cheney (2012) explained, “grooming on Tuesday can create an emotional bond that causes meat sharing on Saturday afternoon” (p. 167).

Similar to humans, nonhuman primates cooperate in a more contingent manner with less bonded partners (de Waal, 1997; Seyfarth & Cheney, 1984, 2012). Most experimental evidence for short-term contingencies comes from cooperation outside of social bonds (see below), which is consistent with the expected difficulty of altering a long-term social bond in a short window of time (Brosnan et al., 2009; Melis, Hare, & Tomasello, 2008). When Melis et al. (2008) found no clear evidence of contingency in two reciprocity experiments with captive chimpanzees, pre-existing social bonds may have been a confounding factor, because one particular chimp would always pull for a specific partner. One lesson here is that experiments on partner control should either use previously unfamiliar subjects or somehow account for the history of past interaction. This is especially important in the absence of partner choice, discussed below.

**Contingency through partner choice.** Although some authors consider partner choice as a specific mechanism for reciprocity (Schino & Aureli, 2009, 2010a,b), reciprocity is typically equated with partner control (increasing and decreasing investment in a single partner) as opposed to partner choice (Noë & Hammerstein, 2001). However, Trivers (1971) recognized partner choice as a form of reciprocity stating that individuals could reciprocate by “decreasing to a minimum the possible exchanges between himself and a subtle cheater and replacing these with exchanges between a new partner or partners. In short, he can switch friends” (p. 47).

Partner choice is particularly relevant when some individuals have greater access to resources or a greater ability to provide services, increasing their value as social partners. In a particularly persuasive demonstration, Fruteau et al. (2009) manipulated the value of low-ranking wild vervet monkeys and observed the response of social partners. A single low-ranking female was given the ability to open a food cache for her entire social group, which led to an immediate spike in her grooming ratio (grooming received minus given). When a second low-ranking female was chosen to be an additional food provider, her grooming ratio spiked as well, and the first provider’s grooming ratio decreased by roughly half (Fruteau et al., 2009).

Sabbatini et al. (2012) conducted tests of passive food sharing (tolerated theft) in capuchin monkey dyads (partner choice absent) and triads (partner choice allowed). In dyadic tests, food shared
from A to B predicted food shared from B to A. In triadic tests, the within-dyad contingency was weaker and food sharing was predicted best by relationship quality, indicating that prior social bonds are more important than recent past sharing within a dyad. When cooperatively nursing female mice are allowed to choose preferred partners rather than non-preferred partners, they achieve higher direct fitness and more egalitarian reproductive outcomes (Koenig, 1994, 2006; Weidt et al., 2008, 2014).

**Experimental evidence for short-term contingency.** Short-term contingency and partner choice are not alternatives; many reciprocity experiments testing short-term contingency use a partner choice design. Rutte and Taborsky (2008) trained rats to pull a lever to deliver food to conspecifics, and found that rats were more likely to pull for partners that previously pulled for them. Anonymous help increased pulling by 20% and help from the same partner increased it an additional 51% (Rutte & Taborsky 2008).

Under natural conditions, short-term contingency should be most obvious in scenarios where partner choice is reduced or absent. Examples include male songbirds on neighboring territories (discussed below under “temporary fitness cost paradox”) or mated pairs raising offspring together. Great Tit parents were found to feed nestlings in a balanced alternating pattern unexplainable by foraging or begging times. Each parent increased feeding rates after their partners contributed, but reduced their feeding rate by about 25% until their partner contributed (Johnstone et al., 2014).

Experimental evidence of reciprocity comes from mobbing behavior of birds. Krams et al. (2008) used fake owls to induce cooperative mobbing in 44 triads of pied flycatcher mated pairs, with each triad consisting of three equidistant nestboxes (A, B, and C). Pair A was exposed to a fake owl near their nestbox to induce mobbing, pair B was held captive and prevented from mobbing, and pair C was left untreated, such that pair C always helped pair A with mobbing, but pair B could not. The authors then simultaneously presented pairs B and C with owls, and tested at which nestbox pair A would choose to help. In 30 of 32 trials, pair A helped pair C. In a follow-up experiment, pair B was presented with an owl. In 8 of 9 trials, pair C, but not pair A, joined B in mobbing, as expected if mobbing efforts are reciprocated in a contingent manner.

Like most claims of reciprocity, this conclusion has been strongly disputed. Russell and Wright (2009) implied that reciprocity was too cognitively difficult for this species (the calculated reciprocity error), and did not consider the form of helping to be costly (see ‘the temporary fitness cost paradox’ below and Wheatcroft & Krams, 2009). Connor (2010) suggested that pair A did not help pair B in order to avoid a potential parasite infestation. These alternative hypothesis seem to assume that reciprocity is highly unlikely a priori.

Krama et al. (2012) ruled out the possibility that reciprocal mobbing at nestboxes was purely a byproduct benefit by showing that the degree of contingency was dependent on the costs and benefits. In the original study, nestboxes were 48-54 m apart. At closer distances (20-24 m apart), they found that subjects always helped neighbors mob regardless of past defections. At farther distances (69-84 m), the original finding was again replicated: pairs helped neighboring pairs contingent on prior help. When the predator was nearby and benefits of mobbing were very high, it was always worth mobbing and any benefit to others was incidental and not enforced. When the predator was farther away, the mobbing was more of a cooperative investment enforced by reciprocity. Hence, reciprocity can involve both byproduct benefits and enforced benefits with their relative importance determined by circumstances.

The degree of reciprocity was also sensitive to whether the failure of partners to mob was caused by their absence (“the excuse principle” Krams et al., 2013). To simulate voluntary defection, the experimenters removed pair B, but played pair B alarm calls to simulate their presence. To simulate involuntary absence, the experimenters completely removed pair B during the predator presentation. When pair B birds appeared present but unwilling to join, pair A only helped pair B in only 2 of 20 cases, but when pair B was completely absent, pair A helped the mob in 20 of 21 cases.

**The Temporary Fitness Cost Paradox**

Clutton-Brock (2009, Table 1) argued that no putative case of reciprocity has demonstrated that “assistance has a net fitness cost at the time it is provided” (p. 54). This is an extremely difficult, if not
impossible, demonstration given that opportunity costs, energetic costs, and increased mortality risk (e.g., predator inspection by small fish: Milinski, Lüthi, Eggler, & Parker, 1997; food sharing in vampire bats: Wilkinson, 1984) have been considered insufficient evidence (Clutton-Brock, 2009; Pusey & Packer, 1997). This temporary fitness costs paradox stems from the fact that many authors including Trivers (1971) define altruism based on short-term payoffs rather than lifetime fitness costs. Although this definition is closer to popular usage, it has led to much confusion in the social evolution literature (West et al., 2007b).

The temporary fitness cost paradox is equivalent to saying that reciprocity assumes that helping others poses a risk, the possibility of cheating, even though the consequence of reciprocity is to prevent cheating. In other words, demonstrating reciprocity requires showing that it doesn’t perform its function. This paradox is rooted in a deeper problem regarding the notion of byproducts and inevitable returns (see ‘the byproduct ambiguity’ below).

One point of the temporary cost requirement is to exclude behaviors that are not forms of helping. For example, several authors have viewed mutual restraint among neighboring male songbirds as reciprocity (Akçay et al., 2009; Axelrod & Hamilton, 1981; Getty, 1987; Godard, 1993; Hyman, 2002). As male songbirds on neighboring territories become familiar they tend to reduce territorial defense and vocal aggression towards one another as compared to strangers (‘the Dear Enemy effect’). Playback studies simulating territorial intrusions by neighboring males found that male hooded warblers increased vocal aggression after playback of those same neighbors compared to control playbacks of other males (Godard, 1993). In similar playback tests, male song sparrows increased their vocal retaliation to previously intruding neighbors but not to others (Akçay et al., 2009). Male red-winged blackbirds did not demonstrate the Dear Enemy effect given that they were more aggressive to neighbors than to strangers, but they did appear to contiguously retaliate against neighbors based on their past actions (Olendorf, Getty, & Scribner, 2004).

Is it fair to call this contingent restraint ‘reciprocity’? Some argue that restraint is not ‘costly’ enough (e.g., Koenig, 1988), but such distinctions are semantic. Fitness costs and benefits are always relative to possible options. When an animal allows only some individuals to use a burrow, feed at a carcass, or take food from its hand, this can be seen equivalently as either conditional punishment or reward. Arguments about whether the Dear Enemy effect should ‘count’ as reciprocity just detract from more important points, for instance, that enforcement of mutual benefit by short-term contingency differs by species, does not require sophisticated cognition, and might be more clear when partner choice is limited by natural circumstances.

The Byproduct Ambiguity

Reciprocity involves mutual enforcement through cooperative investments contingent on cooperative returns. By contrast, ‘pseudoreciprocity’ does not require enforcement because cooperative investments simply enable inevitable byproduct returns (Bergmüller, Johnstone, Russell, & Bshary, 2007; Bshary, 2010; Connor, 1995a, 2010). Pseudoreciprocity assumes that the returns are self-serving byproducts and hence bestowed automatically. Whereas reciprocity involves symmetrical investments, pseudoreciprocity is inherently asymmetrical because it assumes that only one partner makes an investment. Pseudoreciprocity and other byproduct models have been posed as more plausible alternative explanations for almost all putative cases of reciprocity (e.g., Bshary, 2010; Connor, 2010; Raihani & Bshary, 2011). Despite the purportedly clear theoretical distinctions, it is often unclear both how to classify real cases, and why it would be useful to do so. As an illustration, consider one of the most contested claims of reciprocity - predator inspection in fish.

Predator inspection in fish: A case study of byproduct ambiguity. Pairs of fish sometimes approach and inspect larger predatory fish, presumably to assess the situation while maintaining the safety of a companion. The reciprocity explanation claims that fish enforce partner cooperation by approaching closer only if the partner swims beside them (Milinski, 1987). Evidence suggests that predator approach behavior is riskier for both single fish (Pitcher, Green, & Magurran, 1986) and leading fish (Milinski et al.,
Predator inspection involves partner recognition and is contingent on a partner’s past and present predator inspection behavior (Dugatkin, 1988, 1997; Dugatkin & Alfieri, 1991; Milinski, 1987; Milinski, David, & Kettler, 1990), and is more likely to occur with particular partners that have histories of other past social interactions (Croft et al., 2006). Differences in predator inspection behavior of fish from habitats with either high or low predation suggest that the behavior has been shaped by natural selection (Dugatkin & Alfieri, 1992).

Like the similar mobbing behavior in pied flycatchers, this claim of reciprocity has attracted much criticism. One alternative byproduct model argued that the “two individuals jointly adopt the same actions they would perform if alone” (Stephens, Anderson, & Benson, 1997, p. 130), and some authors argued that the movements result even in the absence of a predator (Masters & Waite, 1990; Stephens et al., 1997). Therefore, the supposed cooperation actually resulted from “the simple statistical combination of individual orientation to a predator and attraction to a companion” (Stephens et al., 1997, p. 129). However, other studies using different species present contradicting evidence that inspection is indeed contingent on the existence of a predator (Dugatkin, 1991). Moreover, the observation that fish have preferred inspection partners (Croft et al., 2006; Dugatkin & Alfieri, 1991; Dugatkin, 1997, Milinski et al., 1990) cannot be reconciled with a simple model that assumes no social interactions.

A more nuanced byproduct explanation assumed preference for previously bold individuals, consistent partner choice, and the idea that fish remembering a specific partner “could ‘trust’ it to be bold during subsequent interactions” (Connor, 1996, p. 453). The difference between partner choice for fish that are ‘bold’ versus ‘cooperative’ is admittedly semantic (Connor 1996), and the distinction between this byproduct and reciprocity model is based not on the decisions of the fish but on different interpretations of the costs and benefits. For example, the payoff matrix for leading and lagging behind might not match a Prisoner’s Dilemma but rather a Hawk-Dove Game (also called Snowdrift or Chicken Game), such that bold leaders (dove) do better with other bold leaders but it will still pay to boldly lead with a parasitic laggard (hawk) because two laggards do worst of all (Noë, 2006). According to Connor (1996) and Stephens et al. (1997), this would mean the behavior is not reciprocity. As Stephens et al. (1997) summarized, “the only unambiguous way to distinguish between competing economic models of predator approach is by objective measurement of the economics (i.e. the payoff matrices).” The assumption is that, to understand the behavior, it must be classified as a strategy in a particular game.

The problem is that predator inspection actually looks less like any particular game the more one examines it (Noë, 2006). Fish are not choosing between binary options, such as leading or lagging; rather, they can approach to varying distances at varying speeds. Depending on information about the partner, the actor, and the circumstances, the costs and benefits of leading or lagging can be adjusted continuously by leading ahead a bit less (‘parceling’ Connor, 1995b) or a bit more (‘raising the stakes’ Roberts & Sherratt, 1998). For example, the contingency of mobbing decisions by pied flycatchers varies with predator distance, because the perceived payoffs change with perceived risks (Krama et al., 2012). Views on how well biological reality matches a particular game depend on how literally one takes the game assumptions, how one divides the cooperative behavior into rounds, and how one assigns behavior to the binary choices. For these reasons, debates regarding how well various natural behaviors match the Prisoner’s Dilemma are typically not resolved by additional empirical evidence (Clements & Stephens, 1995; Doebeli & Hauert, 2005; Dugatkin, 1997; Milinski et al., 1997; Noë, 2006; Pusey & Packer, 1997; Raihani & Bshary, 2011; Stephens et al., 1997).

**Game payoffs and the byproduct ambiguity.** Game theory payoff structures and their outcome in evolutionary simulations are drastically altered when allowing any additional element of realism such as kinship, spatial structure, partner switching, communication, long-term relationships, power asymmetries, and continuous variation in the size of cooperative investments (Doebeli & Hauert, 2005; Noë, 2006). Payoffs for partners in the real world might also asymmetric, so each individual or type of individual would in effect be playing a different game. Consider a scenario where some lions can lead the rush to protect a territory from intruders or lag behind and get the benefits of defense without paying the costs (Connor, 2010; Doebeli & Hauert, 2005; Heinsohn & Packer, 1995). Territory defense might be a Hawk-Dove Game for male lions because they can lose all their offspring if ousted by a foreign male.
(leading alone > mutual defection). Whereas for female lions the same scenario might be closer to a Prisoner’s Dilemma (mutual defection > leading alone) because they are likely to sacrifice some, but not all, of their reproductive success if the foreign male gains control. In this case, are male lions performing pseudoreciprocity, while female lions are performing reciprocity?

Strategic adaptive behaviors are always reducible to a combination of very simple decision rules, which are themselves byproducts of other adaptations. So if joint predator inspection in fish is shown to be merely based on a foundation of simple byproduct behaviors, this demonstration of how the contingency works does not refute the idea that decisions of fish are enforced by that contingency. The fact that fish benefit from preferentially choosing bold leaders as partners is already enough contingent aid to help enforce cooperation. Partner choice already assumes that fish are keeping track of their partners’ actions and identity, so why would they not use this information to also guide their actions within dyads?

Byproduct explanations are not favored because they are empirically verified; rather they usually act as null hypotheses. Moreover, they explain behaviors already known to exist, rather than make new predictions of what animals should do to maximize fitness. For example, in a review arguing for the absence of evidence for reciprocity, Clutton-Brock (2009) suggested that simple byproducts or pseudoreciprocity could explain elements of mutualism or manipulation such as:

Regularly associating with dominant individuals, and grooming them repeatedly [in order to] habituate them [and gain] shelter from competition” as well as the “establishment and maintenance of long-term mutualistic relationships...[in which] individuals compete to establish relationships with potential protectors, allies or mates, using a wide range of different forms of affiliative behavior, including close association, grooming, support in competitive interactions, reassurance, and consolation. (p. 55).

In all these cases, the individuals are not reciprocating; rather they merely “modify their behavior to take advantage of the fixed responses of conspecifics.” But if such relationships are completely explained by simple byproduct benefits and do not require enforcement, why then do such complex, long-term social relationships correlate with brain size (Dunbar & Shultz, 2007, 2010)?

In many cases, classifying cases as reciprocity or pseudoreciprocity is more clearly semantic. Reciprocal egg-trading by hermaphroditic fish involves the alternated exchange of valuable eggs for fertilization by the partner (Fischer, 1984; Sella, 1985). The reciprocity explanation has been contested using an alternative byproduct model by Connor (1992) who acknowledges that such egg trading represents a costly investment leading to a costly return and that individuals “parcel those benefits to manipulate each other’s optimal strategy” while also arguing that, “in reciprocity, an individual would realize short-term benefits by cheating on any given interaction. This is not the case in the model presented here” (p. 523). Again, this is a semantic distinction, which depends on how one divides behavior into ‘interactions’. A crucial question is whether the reciprocity hypothesis can ever produce testable predictions that cannot be later explained as consistent with a byproduct explanation.

Byproduct benefits and enforced benefits are not mutually exclusive. Indeed, most enforced benefits likely originated as byproduct benefits, because the frequency of cheating can fluctuate in a population such that a given benefit might be considered ‘a fixed response’ or not, depending on the phenotypes currently in the population. When a cooperative trait goes to fixation, this eliminates the selective pressure for enforcement mechanisms such as contingency. Eventually the trait can become unconditional and hence susceptible again to cheating, which can easily arise again from new variation in the cooperative trait (Foster & Kokko, 2006; Imhof, Fudenberg, & Nowak, 2005).

The line between different kinds of byproduct mutualism and pseudoreciprocity can also be subjective. Raihini & Bshary (2011) explain that seed dispersal is either byproduct mutualism or pseudoreciprocity depending on which organism’s perspective is taken:
The plant invests resources into making seeds that are attractive to some animals. This enables a self-serving response from the animal that eats the seed and later disperses it via defecation. Because the plant makes an initial investment in the interaction, but there is no potential to benefit from reducing this investment, we can explain the plant’s investment with the concept of positive pseudo-reciprocity. The animal, on the other hand, simply eats the seed and later defecates: there is no investment and the benefits to the plant are a by-product of the animal’s own self-serving behavior. (p. 1635)

The authors assume that there is no potential benefit for a plant to reduce its investment, but this is only because the fitness of a plant that produces poor fruits would be reduced by partner choice. If a particular fruit tree provides poor fruit, animal foragers stay away. This is because animals make a costly investment in selecting fruits to open, eat, or carry away. From the plant’s perspective, the cooperative returns (seed dispersal) are thus not fixed, but depend on the size of the investment (fruit quantity and quality). From the animal’s perspective, the cooperative returns (fruit quality) might depend on the cooperative investment (choosing to move to one fruit tree over another). Here, we see that the line between byproducts and enforced benefits is blurred further.

Defining Reciprocity

Evolutionary explanations of cooperation are drawn from several academic sub-fields, leading to many semantic misunderstandings and disagreements (West et al., 2007b). The semantic framework one chooses ultimately depends on what is most useful. But for authors discussing reciprocity, it will be particularly important to define their terms, because reciprocity has many different contradictory meanings in the literature. In this review, I defined reciprocity as occurring when individuals make contingent cooperative investments based on past or expected returns. I believe that this simple, testable definition best captures the original broad concept described by Trivers (1971) and Axelrod and Hamilton (1981). Under this definition, reciprocity is a broad overarching term for conditional enforcement of direct fitness cooperation, including sanctions (Denison, 2000; Kiers et al., 2003; West, Kiers, Simms, & Denison, 2002), reciprocal rewards (Kiers et al., 2011), partner control, and partner choice (Foster & Wenseleers, 2006; Noë & Hammerstein, 2001).

Conclusion

Three key theoretical frameworks have guided empirical studies of cooperation. Inclusive fitness theory (Hamilton, 1964) solved the puzzle of altruism. Reciprocity theory (Axelrod & Hamilton, 1981; Trivers, 1971) illustrated the roles of contingency and frequency-dependent selection in cooperation. Biological market theory (Noë & Hammerstein, 1994) clarified the importance of partner choice and asymmetries in exchange rates. Unfortunately, several unnecessary controversies have resulted from incompatible modeling approaches and semantic frameworks that actually make the same predictions in the real world. One example is the social evolution debate regarding inclusive fitness and multi-level selection (or ‘kin selection versus group selection,’ see Marshall, 2011; West et al. 2007b). Similarly, reciprocity involves a number of competing semantic frameworks. These semantic differences can lead to disagreements about facts, when for example authors mistakenly believe that behaviors described as reciprocity (in a broad sense) are unlikely to be real or important, because reciprocity (in a narrow sense) is rare. Rather than subjectively fitting behaviors to a game metaphor, a broader notion of reciprocity allows researchers to focus on testing the relative importance of different social and ecological factors that influence helping behavior. Hopefully, this review will help distinguish real alternative hypotheses from semantic disagreements based on modeling preferences (“all models are wrong, but some are useful,” Box & Draper, 1987, p. 424).

Cooperative traits cannot always be clearly classified as byproducts versus enforced, direct fitness versus indirect fitness, or altruistic versus mutualistic. Many cooperative behaviors, especially those in complex animal societies, are supported not by a singular mechanism, but rather by a complex interacting
set of decision rules that take into account multiple factors such as genetic relatedness, partner choice, short-term returns, and long-term prior relationships (e.g., cooperative breeding in cichlids: Zöttl, Heg, Chervert, & Taborsky, 2013; food sharing in primates: Jaeggi & Gurven, 2013; Silk, Brosnan, Henrich, Lambeth, & Shapiro, 2013; food sharing in vampire bats: Carter & Wilkinson, 2013a).

How then should we classify various mechanisms (and should we even try to)? One proposal is to avoid the term reciprocity and simply refer to ‘cooperative investments’ and ‘cooperative returns’ (Noë 2006). However, simply abandoning the term ‘reciprocity’ cannot resolve past controversies or clarify connections between recent findings and older studies. In fact, using the terms ‘investment’ and ‘return' already assumes much of what reciprocity predicts— that helping another individual is a conditional investment leading to a return that is not fixed. Whatever this phenomenon is called, it is clearly important across many cooperative organisms with repeating interactions.

Acknowledgements

Jerry Wilkinson, Jennifer Vonk, and two anonymous reviewers provided comments that greatly improved the manuscript. GGC is supported by a Ford Foundation Predoctoral Fellowship administered by the National Academy of Sciences and his work on vampire bats is supported by the National Science Foundation under Grant No. (1311336).

References


