Reciprocity creates a stake in one’s partner, or why you should cooperate even when anonymous

Pat Barclay

Department of Psychology, University of Guelph, 50 Stone Rd. E., Guelph, ON Canada, N1G 2W1

PB, 0000-0002-7905-9069

Why do we care so much for friends, even making sacrifices for them they cannot repay or never know about? When organisms engage in reciprocity, they have a stake in their partner’s survival and wellbeing so the reciprocal relationship can persist. This stake (aka fitness interdependence) makes organisms willing to help beyond the existing reciprocal arrangement (e.g. anonymously). I demonstrate this with two mathematical models in which organisms play a prisoner’s dilemma, and where helping keeps their partner alive and well. Both models shows that reciprocity creates a stake in partners’ welfare: those who help a cooperative partner—even when anonymous—do better than those who do not, because they keep that cooperative partner in good enough condition to continue the reciprocal relationship. ‘Machiavellian’ cooperators, who defect when anonymous, do worse because their partners become incapacitated. This work highlights the fact that reciprocity and stake are not separate evolutionary processes, but are inherently linked.

1. Introduction

Organisms often have a stake in the welfare of other organisms. If organism A does something that happens to benefit organism B, then B has a vested interest in helping A to ensure that A remains alive and well and able to benefit B. For example, imagine a tree that provides useful shade to organisms beneath it. The shaded organisms have a vested interest in protecting the tree and helping it grow, because the benefits of continued shade outweigh the cost of helping the tree. This is not reciprocity and requires no reputation: the tree pays no cost and need not notice the help it receives; it just continues to act in its selfish interest and grow. However, if other organisms benefit from that growth, they have a vested interest in promoting the tree’s survival and growth so it can provide more shade. This principle applies to cooperation within or between species and has been proposed by multiple researchers under many names: stake [1]; pseudo-reciprocity [2]; by-product reciprocity [3]; partnership [4]; group augmentation [5]; interdependence [6,7]; irreplaceability [8]; and vested interests [9]. Despite being proposed multiple times, stake has been studied less than other causes of helping like kinship or reciprocity.

Stake can occur in many different kinds of interactions. Ants protect the ferns and acacias they live in, because those plants provide better homes when un molested by herbivores (e.g. [3]). When hunters share food widely, good hunters who get sick are nursed back to health by others, so the hunters can return to providing food for everyone [10]. Meerkats and possibly owls help groupmates because by doing so, they increase group size and thus decrease their personal predation risk [5,11]. More generally, many species help groupmates in order to increase group size, because more groupmates means more individuals to detect predators, forage for food that can be scrounged, fend off hostile groups or mate with. In particular, organisms have a stake in their monogamous mates: if B will only ever reproduce with A, then B should value A almost as much as itself (excepting where nepotistic interests differ). Furthermore, if A is
raising B’s offspring, then B has an ongoing stake in A’s welfare even after the relationship ends.

In practice, it is often hard to distinguish between reciprocity and stake. If B helps A because A provides benefits to B, then does it matter whether those benefits are actively provided at cost to A (reciprocity) or passively produced by-products of A’s selfish actions (stake, pseudo-reciprocity, by-product reciprocity)? Here I suggest that this distinction does not matter unless it affects the surety of A’s help. When A provides benefits to B—for any reason—this gives B a stake in A’s welfare. As such, B has a vested interest in helping A, so that A can continue to provide those benefits. Furthermore, B will help even if A will never find out about the help, because B benefits from A’s survival and continued ability to help B. Thus, a relationship that starts as reciprocity can develop into one based on stake—reciprocity creates stake.

Some examples can help demonstrate how reciprocity creates stake. Imagine two parents taking turns caring for mutual offspring, in a reciprocal manner. The more that the father invests, the more stake the mother has in his continued wellbeing so that he can continue to invest, and vice versa. If the father defects by stopping his paternal investment, then the mother no longer has any stake in his wellbeing—her stake depends on his reciprocity (and vice versa). Second, imagine two organisms reciprocally exchanging food or coalitional support. If one partner were at risk of injury, the other partner has a vested interest in helping to keep its cooperative partner healthy and continuing to provide food or aid. This underlies need-based transfers towards those who would help you in your need, such as need-based transfers of cattle among the Maasai of Kenya [6,12]. Furthermore, it underlies many long-term partnerships like human friendships [7], which start like superficial reciprocity but deepen into mutual concern for each other’s welfare, and are damaged if one friend does not support the other [8]. Third, imagine two different classes of organisms engaged in reciprocity, like different species or economic roles. If the lone farmer in a town sells her produce to the lone grocer, this is a simple economic exchange of goods for money. However, if the grocer’s shop burns down, then the farmer has a vested interest in rebuilding it—even if the grocer will never know who rebuilt it—so that she has somewhere to sell her produce. Of course, organisms might still want their partners to know, but this willingness to help anonymously is a sure sign of stake and would not occur with reciprocity alone (excepting mistakes [9]).

Little research has examined stake within reciprocal relationships. Some computer simulations have examined risk-pooling via need-based transfers of resources [12–14], but they did not include the possibility of defection. As such, while important, those models cannot distinguish between stake and reciprocity, let alone how they are intertwined and how one creates the other. To show the stability of cooperation, a model needs to compare cooperation against defection.

Here I present two mathematical models to show that having a reciprocity-based relationship with someone creates a stake in that individual’s welfare. Those who help—even when anonymously—do better under many conditions than those who do not help. Just like the ants who protect a host plant from herbivores that the plant need never know about, reciprocal partners might help each other even if the partner never knows about the help.

2. Model A: general model of paying to save a reciprocal partner

Imagine two organisms cooperating reciprocally in a multi-round prisoner’s dilemma. Each round it costs c to confer benefit b upon a partner (b > c). The probability of a future round is w; this probability is independent of how many rounds the pair has already been together. Thus, at the start of any round, the pair is expected to last an average of 1/(1−w) rounds more including the current round (see electronic supplementary material for not including the current round). To a cooperator, having a cooperative partner is worth b−c for each of the expected 1/(1−w) remaining rounds. By contrast, an uncooperative partner provides no benefits, so bad partners are worth nothing to conditional cooperators. An organism without a partner pays no cost of cooperation but also receives no benefit.

Suppose that at some point, one organism is about to die (e.g., predation, starvation) or become incapable of continuing the reciprocal relationship (e.g., injury, incapacitation, bankruptcy, emigration), but it can be saved before the round by its partner at some cost a. It is worth paying that cost to save a cooperative partner whenever (b−c)/(1−w) > a, which can be rewritten as:

\[(b−c) > a(1−w)\].

(inequality 1)

Note that saving a partner is not ‘reciprocated’ by the partner: inequality 1 holds true even if we assume that partners do not know about having been saved and do not change their behaviour as a result. Instead, the benefits of saving a partner are that the partner remains in good condition to continue the existing reciprocal relationship. Even anonymous help can evolve because it preserves existing reciprocal relationships. Furthermore, prior rounds are not strictly necessary: as long as the partner probably will reciprocate in the future and this is somehow inferred (e.g., based on past reciprocation, or if reciprocators are common), then that expected future reciprocation is enough to create a stake in their welfare (see electronic supplementary material, section 1b).

By contrast, an uncooperative partner is not worth keeping alive because they provide no benefits—there are no conditions where \(0 > a(1−w)\) unless a is negative (you’d have to pay me to save you). In fact, organisms may benefit from an uncooperative partner’s demise and may even pay to hasten that demise, if this allows them to find a partner who will reciprocate (see electronic supplementary material, section 1a). Therefore, it’s worthwhile to keep reciprocators alive, but not non-reciprocators. The electronic supplementary material (section 1b) shows that reciprocation is required: reciprocators dominate unconditional cooperators whenever there are any defectors in the population.

Model A is very simple and contains very few assumptions. For example, this model applies whether a, b and c represent costs and benefits in terms of survival or fecundity. It also applies whether the reciprocal relationship would end due to death, incapacitation, emigration (e.g., due to insufficient food) or any other preventable reason. As such, it has broad generality. The few assumptions are examined in electronic supplementary material, such as whether helping and harming occur after a round instead of before (section 2a). In all model variations, there is a wide range of biologically realistic parameters where it pays to save good partners.

What happens when dead or incapacitated partners can be replaced? In the electronic supplementary material
(section 2c), I allow organisms to find a new partner with probability \( f \) each round, and where \( p \) is the probability that one’s new partner is a cooperator. When replacement partners are possible, it pays to keep a partner alive when:

\[
(1 - f)(b - c) - \frac{f(1 - p)(b - cw)}{1 - w(1 - f)} > a. \tag{inequality 2}
\]

Organisms have more stake in their partners when the probability of future rounds \((w)\) is higher, the proportion of cooperators \((p)\) is lower and the gains from cooperation \((b - c)\) are larger. The ease of replacing partners \((f)\) usually reduces one’s stake in one’s current partner, except when cooperators are rare enough that one might pair with a bad replacement—this occurs when \( p < (1 - w)/(b - cw) \) (see electronic supplementary material, section 2c, figures S2 and S3). The probability of finding a new partner can vary between individuals due to partner choice (e.g. [15]): if an individual is a less desirable partner than others (e.g. less attractive, lower status, worse cooperator), then they will experience a lower probability of finding a replacement partner or of having the replacement be a cooperator, and will thus have a greater stake in their current partner.

3. Model B: modelling survival as the currency of cooperation

(a) Basic model: prisoner’s dilemma with survival as the currency

To give a specific example, Model B presents a modified prisoner’s dilemma where the currency is the probability of surviving each round. Each game is divided into \( n \) rounds. Each player survives any given round with baseline probability \( w \), so in the absence of social effects, each player’s probability of surviving until the end of the game is \( w^n \). Each player has a single partner for the game. In any given round, a player can decrease its own survivability by \( c \) to increase its partner’s survivability by \( b \). Thus, a defector survives with probability \( T = w + b \) against a cooperator and \( P = w \) against another defector, whereas a cooperator survives with probability \( R = w + b - c \) against a cooperator and \( S = w - c \) against a defector.\(^1\) Dead players do not interact: each player only pays costs or receives benefits from its partner in round \( t \) if its partner has survived the previous \( t - 1 \) rounds. After \( n \) rounds, all surviving players breed equally. Thus, each player’s fitness is proportional to its probability of surviving the \( n \) rounds, and it maximizes fitness by maximizing its probability of surviving the \( n \) rounds.

A player’s probability of surviving \( n \) rounds (its ‘payoff’) is the product of its survival probability in each of \( n \) rounds. Let \( i \) be the payoff of strategy \( i \) playing against strategy \( j \). When a defector plays a defector (henceforth ‘AllD’, or ‘D’), it neither pays costs nor receives benefits, so its payoff against itself is:

\[
D_D = \prod_{t=1}^{n} w. \tag{3.1}
\]

For cooperators, I use a version of tit-for-tat (TFT), which cooperates on the first round and thereafter imitates its partner’s previous action, even when anonymous (see §3b). TFT-like strategies are good proxies for all conditional cooperators because TFT is simple, well studied and easy to model (e.g. [16]); the same principles apply to more complex conditional cooperators. When TFT plays another TFT, they both cooperate in all rounds, so TFT earns \( w \) plus \( b - c \) in any round its partner is alive. Therefore, its payoff against itself is

\[
T_T = \prod_{t=1}^{n} (w + (b - c)(w + b - c)^{t-1}). \tag{3.2}
\]

In the first round that TFT faces AllD, TFT pays a cost (payoff of \( w - c \)) and AllD receives a benefit (payoff of \( w + b \)). After that round, they both defect on each other in subsequent rounds if AllD’s defection is observed (see below).

(b) Anonymous cooperation and defection

In other studies of prisoner’s dilemmas, after each round both players find out what their partners did, such that those actions can influence subsequent decisions. I relax this assumption: I let proportion \( x \) of rounds be observed (i.e. players find out their partners’ actions), whereas the other \( 1 - x \) rounds are anonymous (i.e. neither partner discovers what its partner did, such that those actions cannot influence subsequent decisions). I assume that players know which rounds are anonymous, but do not know their partners’ actions those rounds. Neither AllD nor TFT change their behaviour under observation or anonymity, but it affects whether TFT discovers AllD’s defection. AllD’s deflections remain undiscovered by TFT if all previous \( t \) rounds were anonymous (probability \( 1 - x \) each round); if so then TFT continues to cooperate until it observes defection.\(^2\) Thus, AllD’s and TFT’s payoffs against each other are

\[
D_T = \prod_{t=1}^{n} (w + b(1 - x)^{t-1}(w - c)^{t-1}) \tag{3.3}
\]

and

\[
T_D = \prod_{t=1}^{n} (w - c(1 - x)^{t-1}(w + b)^{t-1}). \tag{3.4}
\]

I introduce another TFT-like strategy, Machiavelli (M), who acts like TFT except that it defects in all anonymous rounds. By acting like TFT when observed, Machiavelli gets the benefit of long-term cooperation with conditional cooperators like TFT. It also gets the benefit of defecting when anonymous (i.e. it only pays cost \( c \) in \( x \) rounds), and it is never discovered cheating because no one’s actions in those rounds are ever known. However, because Machiavelli defects when anonymous, its cooperative partners are more likely to die in anonymous rounds (i.e. its partners only receive benefit \( b \) in \( x \) rounds). Pairings involving a Machiavellian player (M) have the following payoffs:

\[
D_M = \prod_{t=1}^{n} (w + bx(1 - x)^{t-1}w^{t-1}), \tag{3.5}
\]

\[
M_D = \prod_{t=1}^{n} (w - cx(1 - x)^{t-1}w^{t-1}), \tag{3.6}
\]

\[
T_M = \prod_{t=1}^{n} (w + (bx - c)(x(w + b - c) + (1 - x)(w + b))^{t-1}), \tag{3.7}
\]

\[
M_T = \prod_{t=1}^{n} (w + (bx - c)(x(w + b - c) + (1 - x)(w - c))^{t-1}), \tag{3.8}
\]

and

\[
M_M = \prod_{t=1}^{n} (w + x(b - c)(x(w + b - c) + (1 - x)x)^{t-1}). \tag{3.9}
\]
The only difference between TFT and Machiavelli is that Machiavelli defects when anonymous. Therefore, it pays to cooperate—even when anonymous—whenever the payoff to TFT exceeds the payoff to Machiavelli. The full equations of comparative payoffs (e.g. $T_I$ versus $D_I$) are unwieldy because of the product terms, so I present the information graphically below and with a wider range of parameters in electronic supplementary material, section 5. Machiavelli is indistinguishable from TFT when $x = 1$ and indistinguishable from AllD when $x = 0$, but to avoid undefined numbers associated with zeros, I restrict the range of $x$ from marginally greater than 0 to marginally less than 1.

(c) Results and discussion of model B

(i) No stake: playing against AllD

If one is paired with a defector (AllD), it pays best to always defect, pays worse to play Machiavelli and pays worst of all to play TFT and attempt cooperation even when anonymous: $D_D \geq M_D \geq T_D$ for all parameter values. Thus, in this model, there is no stake without reciprocity.

(ii) Stake: playing against conditional cooperators (TFT) and sneaky Machiavellians (Mach)

If one is paired with some type of cooperator, then it often pays to cooperate (TFT > AllD), even when anonymous (TFT > Mach). Figure 1 shows that AllD pays best at low $b$ (i.e. low benefits to cooperation), Machiavelli pays best at intermediate $b$, and TFT pays best at high $b$. In other words, if there are large gains from cooperation, then it pays to cooperate—even when anonymous—to keep your partner alive and able to provide you with those gains. Furthermore, this result holds whether one is paired with an ‘honest’ cooperator like TFT (figure 1a) or a sneaky Machiavellian who defects when anonymous (figure 1b); as long as one’s partner cooperates some of the time, then it is worth keeping them alive, even if one must do so anonymously.

Figure 1 shows that high observability ($x$) helps TFT outcompete AllD, because partners detect AllD’s defections earlier. In particular, high observability helps TFT when it’s paired with a Machiavellian partner (figure 1b): it only pays to keep a sneaky partner alive if they’re rarely anonymous. See the electronic supplementary material (section 4) for a discussion of how observability helps or hinders Mach against AllD.

The electronic supplementary material broadens the range of parameters to show that it pays to cooperate—anonymously or not—when there are lower costs ($c$) and more rounds ($n$), as in previous models of reciprocity (electronic supplementary material, section 5, figures S5–S8); i.e. stake is higher in longer interactions and when cooperation costs less. Furthermore, a high baseline survivability ($w$) slightly increases the payoffs for cooperation because it increases the likelihood of another round (electronic supplementary material figures S5–S8). Stake can develop even with very few rounds (electronic supplementary material, figures S7 and S8), or even just two (electronic supplementary material, figures S9 and S10); the ‘shadow of the future’ need not be long. Given that TFT often pays better than either other strategy against Machiavelli, this means that TFT can often invade a population of Machiavellians (see critical thresholds and mixed populations in electronic supplementary material, figures S11 and S12).

4. General discussion

When organisms have a stake in another’s survival, they will help even if that recipient remains unaware of that help—a living (but unaware) partner is better than a dead partner. In both models, helping good partners paid off. Model A shows the conditions when it pays to save a reciprocal
partner, and model B shows that TFT (who helps partners even when anonymous) often had a higher payoff with cooperators than did Machiavelli (who cooperated when observed but defected when anonymous). Thus, both models show that reciprocal relationships create a stake in each other’s welfare. Conversely, helping bad partners never paid off, especially not anonymously. As such, it is specifically the reciprocity which creates stake, not some unique feature of the survival-based prisoner’s dilemma in model B. Thus, reciprocal relationships are a source of stake, just like the other causes of stake in the literature (e.g. group augmentation [5]; ants protecting their acacia habitat [3]). Once established, organisms may treat reciprocal benefits the same as by-product benefits (except for their certainty), in that both kinds of benefits are worth preserving, investing in and competing over (e.g. [15,17]).

These models help explain why people develop deep emotional bonds with close friends (e.g. [7]). Friendships may start as reciprocity, but as the reciprocity deepens, we develop a stake in our friends’ welfare. Eventually we value friends for their own sake and are willing to make large sacrifices for them, even sacrifices they can’t repay—our own welfare depends on them being well enough to continue the reciprocal relationship. We may even help under apparently hopeless circumstances if our stake is high enough (e.g. spouse)—high gains are worth gambling against long odds. If someone has not yet helped, but you know they are likely to help when you need it, then you still have a stake in their welfare and should be willing to help them pre-emptively (e.g. ‘osotua’ need-based transfers between Maasai herders [6,12]; see electronic supplementary material, section 1b).

I modelled direct reciprocity, but the results may also apply to indirect reciprocity or broader social networks. Each organism has a stake in whoever might help it, and in whoever helps those who help it, such as others in a network of indirect reciprocity or in a group with common goals (e.g. mutual defence). This generalized stake might explain humans’ willingness to help others in anonymous experiments (though see also mistakes, [9]).

As powerful as reciprocity-based stake is, it has limitations. First, your stake in reciprocal partners depends on how easily you can replace them, and how good the replacements are (see electronic supplementary material, sections 2b and 2c). If good partners are easily replaced, then it can cost less to find a new partner than to save the existing one; this results in less stake and less anonymous helping. Some individuals are more desirable partners and can thus find new partners more easily (e.g. see [15,18]): such individuals have less stake in their current partners. Organisms may strive to make themselves irreplaceable, to ensure that they are valued by their partners [8]. If relationships take time to build up to higher expectation (e.g. ‘raise-the-stakes’ reciprocity [19]), this results in a high stake in one’s partners, because restarting with a new partner means missing out on high cooperation until that relationship matures.

A second limitation is that reciprocity-based stake makes it less useful to punish cheaters, because punishment reduces the partner’s wellbeing [20] and thus the probability that the partner can cooperate in the future. While organisms have no stake in full-time defectors, they do have some stake in part-time defectors or ‘subtle cheaters’ [21] based on their occasional cooperation. One solution is to start with a warning—a small or inconsequential punishment which escalates if defection continues (e.g. [22]). Not only does this warn the defector about future greater punishment, it also signals that if defection continues, the focal agent would have less stake in the defector’s wellbeing (and therefore have less compunction about reducing the defector’s wellbeing).

A third limitation is that while reciprocity creates a stake in a partner’s survival and growth, it does not create a stake in their reproduction. Thus, reciprocal allies might be willing to help their partners to survive and grow, but not to reproduce. In extreme cases, one party may benefit from diverting its partner’s efforts away from reproductive effort towards somatic effort (see electronic supplementary material, section 3), like the ants who castrate their symbiotic plant hosts so the plant invests more in the resources that benefit the ants [23]. Organisms will only have a stake in their partner’s reproduction if they benefit from larger group size (see group augmentation [5]) or if the reciprocal relationships persist across generations, like vertically transmitted mutualisms. Vertically transmitted partnerships will also allow reciprocity-based stake to continue even when agents are older and approaching death.

Despite these limitations, reciprocity-based stake gives organisms a reason to help their partners, even when anonymous, especially when the relationship cannot be replaced immediately. Altogether, rather than being separate forces in the evolution of cooperation, reciprocity is just one of many ways in which organisms have a stake in the wellbeing of their mutualistic partners.

Data accessibility. This article has no additional data.

Competing interests. I declare I have no competing interests.

Funding. This study was funded by the Social Sciences and Humanities Research Council of Canada (SSHRC) grant no. 430287.

Acknowledgements. I thank Stuart West for discussions, Miguel dos Santos for mathematical help, three anonymous reviewers for useful feedback, and the Social Sciences and Humanities Research Council of Canada (SSHRC) for funding (grant no. 430287).

Endnotes

1 For simplicity I assume (for now) linear costs and benefits, with constraint 0 ≤ w − c < w + b ≤ 1 to avoid survival values less than 0 or greater than 0. The electronic supplementary material (section 7) provides a robustness check by examining costs and benefits as a function of residual mortality and survivability.

2 Organisms can only respond to actions they observe (using any sense). Because TFT is normally not defined with respect to unobserved rewards, some readers may wish to call my version ‘TFT-like’ or ‘tit-for-observed-tat (TFOT)’ instead of TFT; I invite them to substitute these terms throughout.

3 I thank an anonymous reviewer for this point.

References


