Supplementary Material for:
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
Phone: 1-519-824-4120 ext. 58247. Fax: 1-519-837-8629.
barclayp@uoguelph.ca, www.patbarclay.com, ORCID 0000-0002-7905-9069

Abstract
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 (a.k.a. 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 Prisoners 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.
**Keywords: Pseudo-reciprocity; fitness interdependence; mutualism; helping; altruism; Prisoner's Dilemma**

**Table of Contents for Supplementary Material:**

## 1) Elaboration on Model A (paying to save a partner)

The main text describes Model A as two organisms playing a Prisoners Dilemma, where each round it costs $c$ to confer benefit $b$ upon one's partner, with probability $w$ of there being another round. At some point, the focal individual's partner faces incapacitation (e.g., death, injury, bankruptcy, emigration due to poor conditions), which the focal individual can prevent by paying some cost $a$. Inequality S1 reprints the conditions when it is worthwhile to pay that cost to save a good reciprocator (from the main text Inequality 1). Saving a good partner is worthwhile when:

$$(b\text{-}c) > a(1\text{-}w) \qquad \text{(Inequality S1)}$$

### 1a) Paying to hasten a bad partner's demise

Unlike a good reciprocator, an uncooperative partner is not worth keeping alive because they provide no benefits. In fact, unconditional cooperators actually benefit from an uncooperative partner's demise, because unconditional cooperators pay $c$ each round their uncooperative partner is alive. As such, unconditional cooperators will pay $h$ to cause an uncooperative partner's demise (or otherwise leave them) whenever $c/(1\text{-}w) > h$, which can be rewritten:

$$c > h(1\text{-}w) \qquad \text{(Inequality S2)}$$

### 1b) Including frequencies of different strategies: effects of reciprocity & uncertainty

Imagine four different strategies for playing the Prisoners Dilemma and paying to save a partner. Unconditional cooperators (C) cooperate in each round of the Prisoners Dilemma and pay to save their partners. Defectors (D) defect in each round of the Prisoners Dilemma and do not pay to save their partners. TFT-Save (TS) and TFT-NotSave (TN) are both Tit-for-Tat players who cooperate on the first round and thereafter imitate their partner, such that they both cooperate with those who cooperate (C, TS, & TN), and defect with defectors after the first round. However, whereas TFT-Save acts like it has a stake in good partners and pays to save those who cooperate (and only those who cooperate), TFT-NotSave acts like it has no stake in anyone and does not pay to save anyone. Let $p_C$, $p_S$, $p_N$ and $p_D$ be the proportion of the population who are unconditional cooperators, TFT-Save, TFT-NotSave, and defectors, respectively.

In the $n$ rounds before a partner needs saving ($n \geq 1$), this is a standard Prisoners Dilemma with unconditional cooperators, defectors, and Tit-for-Tat (TFT-Save and TFT-NotSave are identical up to the point of saving). Thus, if the game ends before $n$ rounds, then there is no saving decision to model, and the analysis is identical to any other Prisoners Dilemma (and thus not requiring analysis here). Thus, I limit my analysis to games that last at least $n$ rounds, such that the focal individual's partner needs to be saved. After the round with saving, the interaction continues with probability $w$. As long as $n \geq 1$, the payoffs for unconditional cooperators, TFT-Save, TFT-NotSave, and defectors are as follows:

$$w_C = (p_C + p_T + p_M)\left(n(b-c) - a + \frac{b-c}{1-w}\right) + p_D\left(-nc - \frac{c}{1-w} - a\right) \qquad \text{(Equation S1)}$$

$$w_{TS} = (p_C + p_T + p_M)\left(n(b-c) - a + \frac{b-c}{1-w}\right) + p_D(-c) \qquad \text{(Equation S2)}$$

$$w_{TN} = (p_C + p_T + p_M)(n(b-c)) + p_D(-c) \qquad \text{(Equation S3)}$$

$$w_D = p_C(nb) + p_D(0) + (p_T + p_M)b \qquad \text{(Equation S4)}$$

In many ways this is like a standard Prisoners Dilemma with TFT, AllC and AllD, which has already been analyzed many times elsewhere and is therefore not worth analyzing again in depth

here. However, three things are noteworthy with these payoffs. First, TFT-Save outperforms TFT-NotSave whenever $(b\text{-}c)/(1\text{-}w)>a$, which corresponds to Inequality 1 in the main text (and Inequality S1 in Supplementary). These are the conditions under which organisms have a stake in good cooperators (C, TS, & TN).

Second, whenever $p_D > 0$, TFT-Save strictly dominates unconditional cooperation. This holds true both for the rounds before and after the saving: unconditional cooperators continue to get suckered by the defectors they save, whereas TFT-Save does not save the defector and thus doesn't continue to get suckered. Thus, conditional cooperation is required for this type of stake to work – it doesn't pay to act like you have a stake in everyone. We can also imagine a version of TFT-Save who *does* save defectors – this version would do worse than one who doesn't pay the cost of saving someone they won't cooperate with.

Third, the model does not require *prior* cooperation in order to have a stake in a partner, just a statistical expectation of *future* reciprocity. Suppose that $n$=0, such that there are no rounds before the focal individual's partner needs saving. Defectors and TFT-NotSave do not save their partners, and thus have no interaction when they are the focal individual. When $n$=0, the payoffs are:

$$w_C = (1 - p_D)\left(\frac{b-c}{1-w} - a\right) + p_D\left(-\frac{c}{1-w} - a\right) \quad \text{(Equation S5)}$$

$$w_{TS} = (1 - p_D)\left(\frac{b-c}{1-w} - a\right) + p_D(-c - a) \quad \text{(Equation S6)}$$

$$w_{TN} = 0 \quad \text{(Equation S7)}$$

$$w_D = 0 \quad \text{(Equation S8)}$$

Once again, TFT-Save dominates unconditional cooperation whenever $p_D>0$, such that reciprocators (TFT-Save) have more stake in saving an unknown partner than do unconditional cooperators (C). Thus, there is more stake with reciprocity than without. Furthermore, TFT-Save outperforms TFT-NoSave (and Defect) whenever:

$$(1 - p_D)\left(\frac{b-c}{1-w} - a\right) > p_D(c + a) \quad \text{(Inequality S3)}$$

This means that as long as defectors are not too common ($p_D$), then the gains from the expected long-term cooperation $(b\text{-}c)/(1\text{-}w)$ outweigh the cost of saving partners ($a$) and the risk of getting suckered by a defector ($c$). Thus, one can have a stake in partners of uncertain cooperativeness, as long there is a reasonable expectation of *future* reciprocity.

## 2) Model A: Variations to test assumptions

Model A is very general. Nevertheless, it contains a few assumptions. Here, I examine two assumptions that it does make: a) the timing of helping and harming; b) whether incapacitated partners can be replaced by other cooperators; and c) whether incapacitated partners can be replaced by cooperators or defectors.

### 2a) Helping and harming that occurs *after* the round instead of before the round

The inequalities in main text (Inequality 1) and in Supplementary (Inequalities S1-S3) assume that helping and harming occur at the start of a round, which then takes place with certainty. The probability of a future round, $w$, only applies to rounds after that. Under this assumption, the payoffs for saving a good reciprocator are $(b\text{-}c) + w(b\text{-}c) + w^2(b\text{-}c) + w^3(b\text{-}c) + \ldots$ to infinity, i.e., $(b\text{-}c)/(1\text{-}w)$ because there are $1/(1\text{-}w)$ expected future rounds.

However, if helping and harming occur *after* a round, it means that there is no future interaction guaranteed because the interaction could end before the next round starts (with probability $1\text{-}w$). Under this assumption, there are only $1/(1\text{-}w)$ -1 expected future rounds. In this case, the expected payoff for saving a cooperative partner is $(b\text{-}c)/(1\text{-}w) - (b\text{-}c)$, such that it is worthwhile to save a partner when:

$$w(b\text{-}c) > a(1\text{-}w) \qquad \text{(Inequality S4)}$$

Similarly, paying $h$ to hasten an uncooperative partner's demise is still worthwhile when:

$$w(b\text{-}c) > h(1\text{-}w) \qquad \text{(Inequality S5)}$$

These inequalities are identical to Inequalities S1 and S2 except for the addition of $w$ to the left-hand term. Thus, the general conclusions about stake in main text still hold, and the conditions for stake are only slightly reduced. Figure S1 compares the two versions of the model when the cost is normalized at $a$=1 (for easy comparison of $b$ and $c$).

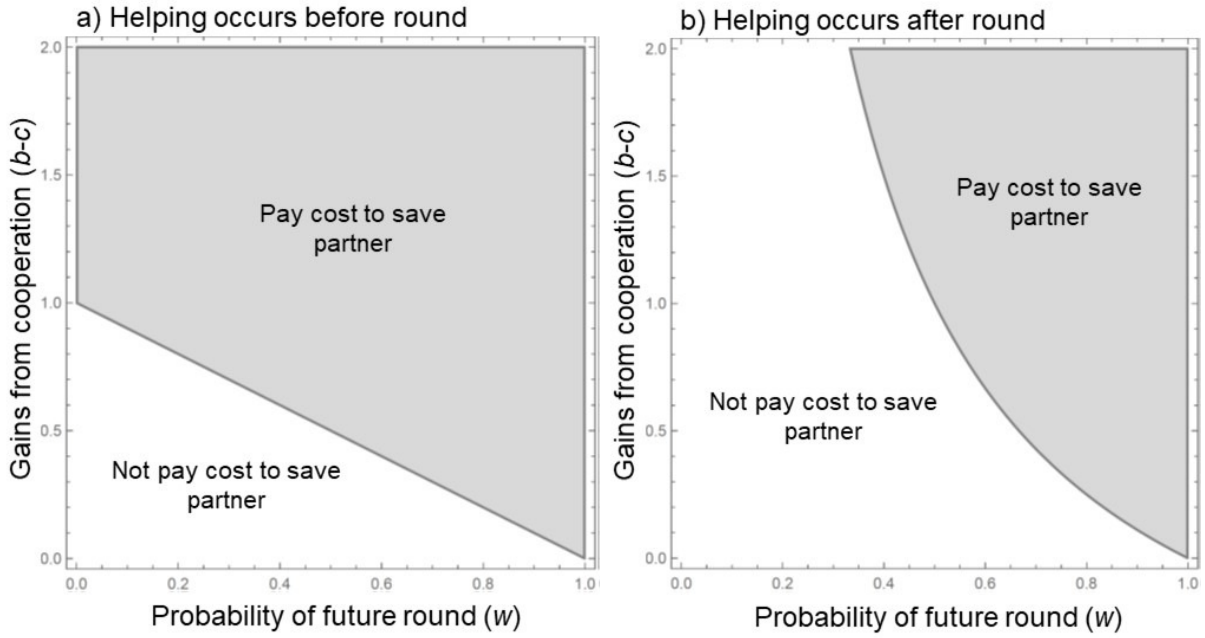


Figure S1: Regions where it is worthwhile to pay cost $a$=1 to save a cooperative partner (grey) versus not save them (white), depending on whether: a) helping takes places immediately before

a round which then occurs with certainty (model in main text); or b) helping takes place after a round such that the next round fails to occur with probability 1-*w* (model in Supplementary 1c).

**2b) When incapacitated partners can be replaced with other cooperators**

What happens if dead, incapacitated, or otherwise unavailable partners can be replaced? For this, I compare the payoff of paired versus unpaired organisms (e.g., those with capable vs. incapable partners) to see if the opportunity cost of lost cooperation outweighs the cost of keeping a partner alive and in good condition. Suppose that each round, an unpaired organism finds a partner with probability *f* and remains solo with probability 1-*f*. Thus, an organism whose partner dies before the round has a probability (1-*f*) of being unpaired in a round where it *would have been paired had its partner survived*. The next round, it has $w(1-f)^2$ probability of still being unpaired in a round where it *would have been paired had its partner survived*. The third round has a $w^2(1-f)^3$ chance, and so on, such that it can expect to be alone for an average of $(1-f)/(1-w(1-f))$ rounds where it would have been paired. (Note: this section assumes that all new partners are cooperators, for example due to assortment including partner choice, see Barclay 2013. The next section relaxes that assumption and allows for replacement by defectors.)

In each round where it *could have had* a living partner who cooperates, it earns *b-c* less than it would have earned *had its partner been alive*. Thus, the benefits of keeping a cooperative partner alive are $\frac{(1-f)(b-c)}{(1-w(1-f))}$. Thus, it is worthwhile to pay *a* to keep a cooperative partner alive whenever:

$$\frac{(1-f)(b-c)}{(1-w(1-f))} > a \qquad \text{(Inequality S6)}$$

Which can be rewritten as:

$$(1-f)(b-c) > a(1-w(1-f)) \qquad \text{(Inequality S7)}$$

In other words, it pays to keep a cooperative partner alive when there are large gains from cooperation (*b-c*), a high probability of future rounds (*w*), and when new partners are hard to find (i.e., low *f*). Note that when *f*=0, Inequalities S6 and S7 simplify to Inequality 1 in the main text.

**2c: When incapacitated partners can be replaced (including by defectors)**

If a dead or incapacitated cooperator can be replaced, but the new partner could be a defector, then it is often better to have kept the cooperator alive and in good condition. To show this, I compare the cost that a focal agent pays to keep a cooperator alive versus the opportunity cost of not being paired with a cooperator. I will use "die" as a shorthand for a partner becoming incapable of helping, whether this occurs due to death, injury, incapacitation, bankruptcy, unavailability, emigration due to poor local conditions, or any other cause.

Let *f* be the probability of finding any replacement partner in any given round. Let *p* be the probability that one's new partner will be a cooperator – this can represent either the proportion of cooperators in the population or one's personal probability of assorting with a cooperator based on how desirable a partner one is (i.e., one's market value, see Barclay, 2013). Defectors (1-*p*) sucker the focal agent for one round (i.e., impose cost *c*) before the focal agent begins reciprocating the defection (i.e., both earn 0). In each round, agents earn 0 if unpaired, *b-c* if paired with a cooperator, -c in their first round paired with a defector, and 0 in any subsequent rounds with that defector. Thus, the *opportunity cost* of letting a cooperative partner die is *b-c* for each round that one remains unpaired, 0 in any round after one finds a new cooperative partner, *b* for the first round that one is paired with a defector, and *b-c* for any subsequent rounds paired

with a defector. Because *w* is independent from round to round, at any given point, any new partnership can be expected to last as long from that point on as would an existing partnership.

How many rounds of cooperation does an agent lose if its partner dies, and at what opportunity cost? The answer is the sum of two infinite series, each representing different opportunity costs across rounds: the probability of still being alone in each round (opportunity cost *b*-*c*); and the probability of meeting a defector *in that round*. If the focal agent meets a defector in a given round, it experiences an opportunity cost of *b* in that round, plus an opportunity cost of *b*-*c* in each of the future rounds expected with that partner. Once it pairs with a defector, it can expect the same number of future rounds with that defector as it could have with its original partner from that point onwards – the expected number of rounds is always 1/(1-*w*) at any given point (including the present round), regardless of previous rounds, because *w* is independent from round to round. Thus, once the focal agent meets a defector, it pays opportunity cost *b* in that round and *b*-*c* in each of the 1/(1-*w*) – 1 expected future rounds. (Below I list these future interactions in the round in which they start, because then it's easier for readers to see the pattern.) If the focal agent finds a cooperator, then it experiences no further opportunity costs because it has successfully replaced its partner – all rounds with a cooperator drop out of the equations for opportunity costs, because the opportunity cost is zero (and will not be mentioned further). Thus, the payoffs for the first five rounds are:

1. $(1-f)(b-c)+f(1-p)\left(b+\left(\frac{1}{1-w}-1\right)(b-c)\right)$
2. $w(1-f)^2(b-c)+w(1-f)f(1-p)\left(b+\left(\frac{1}{1-w}-1\right)(b-c)\right)$
3. $w^2(1-f)^3(b-c)+w^2(1-f)^2f(1-p)\left(b+\left(\frac{1}{1-w}-1\right)(b-c)\right)$
4. $w^3(1-f)^4(b-c)+w^3(1-f)^3f(1-p)\left(b+\left(\frac{1}{1-w}-1\right)(b-c)\right)$
5. $w^4(1-f)^5(b-c)+w^4(1-f)^4f(1-p)\left(b+\left(\frac{1}{1-w}-1\right)(b-c)\right)$

And so on to infinity. The first term in each round (i.e., leftmost term) represents the probability of still being alone that round (i.e., $(1-f)^n$) in a round when one's previous partner would have still been alive (i.e., $w^{n-1}$). The second term in each round represents the probability of having been alone in each previous round and finding a defector *in that round* (i.e., $(1-f)^{n-1}f(1-p)$), in a round where one's previous partner would still have been alive (i.e., $w^{n-1}$). Again, please note that I have included "future rounds with defector" ($b+\left(\frac{1}{1-w}-1\right)(b-c)$) in the round in which the agent first meets the defector, because then it is easier for readers to see the pattern; this only occurs if the agent meets the defector *in that round*.

If we solve the infinite series on the left, we get:

$$\text{Opportunity cost of being alone} = \frac{(1-f)(b-c)}{1-w(1-f)} \qquad \text{(Equation S9)}$$

If we solve the infinite series on the right, we get:

$$\text{Opportunity cost of meeting a defector} = \frac{f(1-p)\left(b+\left(\frac{1}{1-w}-1\right)(b-c)\right)}{1-w(1-f)} \qquad \text{(Equation S10)}$$

The latter can be rewritten as:

$$\text{Opportunity cost of meeting a defector} = \frac{f(1-p)(b-cw)}{(1-w)(1-w(1-f))} \qquad \text{(Equation S11)}$$

Adding up all both types of opportunity costs, it pays to keep a partner alive when:

$$\frac{(1-f)(b-c)}{1-w(1-f)} + \frac{f(1-p)(b-cw)}{(1-w)(1-w(1-f))} > a \qquad \text{(Inequality S8)}$$

In other words, it pays to keep a cooperative partner alive when there are large gains from cooperation ($b$) at low costs ($c$), a high probability of future rounds ($w$), and when cooperators are rare (low $p$). Note that when $f$=0 (i.e., dead partners can never be replaced), Inequality S8 simplifies to Inequality 1 in the main text. When $p$=1 (i.e., all new partners are cooperators), Inequality S8 simplifies to Inequality S6. Inequality S8 can also be rewritten as:

$$\frac{b(1-w-fp-fw)-c(1-w-f+f(2-p)w)}{(1-w)(1-w(1-f))} > a \qquad \text{(Inequality S9)}$$

If it's easy to find a new partner each round (high $f$), does this make it better or worse to save one's existing partner? It depends on whether that new partner will be a cooperator ($p$) or a defector (1-$p$). When $p < c(1-w)/(b-cw)$, too few of the potential new partners are cooperators, so it's better to be alone than to find a new partner who might be a defector. In this case, high encounter rates (high $f$) make it more worthwhile to save existing partners. By contrast, when $p > c(1-w)/(b-cw)$, enough new partners are cooperators, so higher encounter rates (high $f$) mean that one can quickly replace a dead cooperator with a new one. In that case, high encounter rates make it less worthwhile to save existing partners, i.e., less stake in existing partners.

Figure S2 displays Inequalities S8 & S9 graphically and shows when it is worthwhile to pay a cost of $a$=1 to save a partner at the sample parameters of $c$=1 and $b$=2. If one is unlikely to find a new partner immediately, either because it is hard to find partners each round (low $f$) or because the new partner is unlikely to be a cooperator (low $p$), then it is worthwhile to save a partner (i.e., one has a stake in their welfare) even if one does not expect many more rounds with them (low $w$). For context: agents can expect to remain unpaired for (1-$f$)/(1-(1-$f$)) = 1/$f$ -1 rounds on average, such that $f$=0.50 corresponds to one round unpaired, $f$=0.25 corresponds to three rounds unpaired, and $f$=0.10 corresponds to nine rounds unpaired, on average. Thus, if agents often miss at least one round before finding a new partner, then it means $f$ is less than 0.5, because $f$>0.5 corresponds to spending less than a single round unpaired on average.

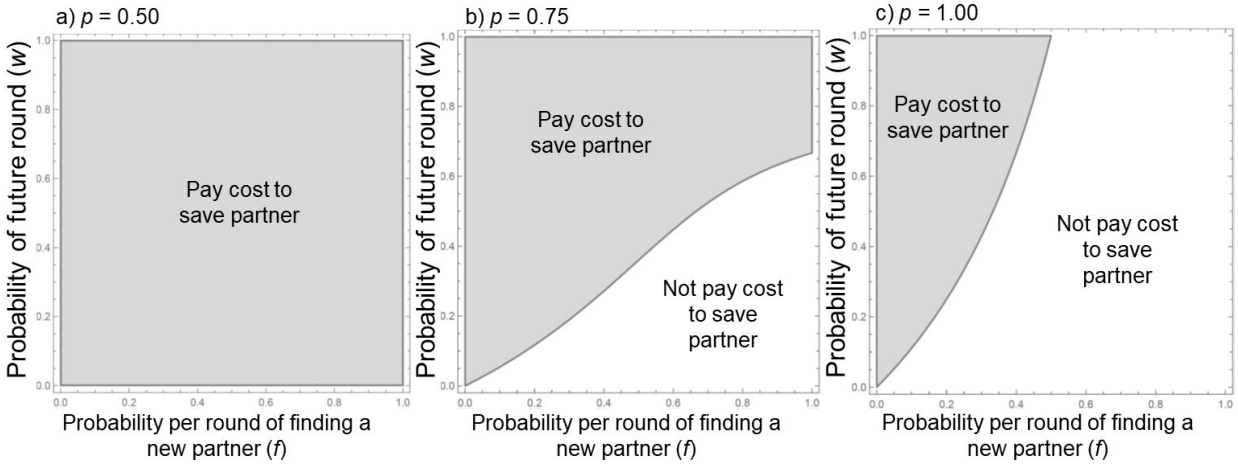


Figure S2: Regions where it is worthwhile to pay cost $a$=1 to save a cooperative partner (grey) versus not save them (white), depending on whether the probability of a new partner being a cooperator is: a) $p$=0.50; b) $p$=0.75; c) $p$=1.00. Parameters displayed: $c$=1, $b$=2.

How much *would* someone be willing to pay to save a partner, i.e., how *much* stake do they have? Figure S3 shows the cost ($a$) that one would be willing to expend to save a partner, for a representative benefit/cost ratio of $c$=1 and $b$=2. Agents are willing to expend much more than the gains from a single round when there is a high probability of future rounds ($w$) and either a low probability of finding a partner each round (low $f$) or a reasonable chance that a new partner is a defector (e.g., $p$<0.75). When the benefits of cooperation ($b$) are higher (e.g., b=5), then agents are willing to pay much more to save partners, even at much lower $w$ and higher $f$ and $p$ (results not shown, available upon request).

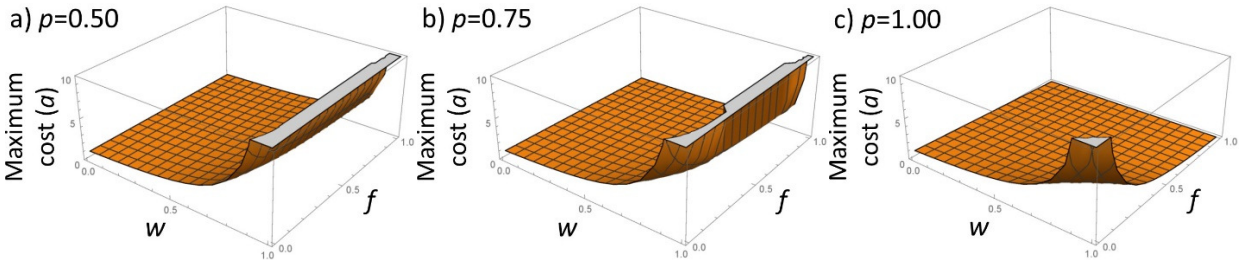


Figure S3: Maximum cost ($a$) that is worth paying to keep a partner alive when $c$=1 and $b$=2, depending on the probability of each future round ($w$), the probability of finding another partner each round ($f$), and the probability that a new partner is a cooperator ($p$), at: a) $p$=0.50; b) $p$=0.75; c) $p$=1.00. Values of $a$ beyond 10 are not displayed – the flat grey area represents values above 10. Note that the value of maximum $a$ is above 1 (i.e., the cost of cooperation $c$) for most parameter values in this graph, unless $p$ and $f$ are both very high (recall that $f$>0.50 corresponds to spending less than a single round unpaired on average). At high $w$ and either low $f$ or low $p$, the maximum $a$ can be several times higher than either $c$ or $b$. So while the probability of replacing a partner does undermine stake, there can still be considerable stake at high $w$.

The probability of finding a good partner can vary between individuals. For example, imagine that there is some assortment in the population. Imagine an individual who is an undesirable partner (e.g., less attractive, less reliable cooperator, lower status) and is thus unlikely to attract another good cooperator (see Barclay, 2013). That low-market value individual has more stake in their current partners because they're either less likely to pair at all or less likely to attract another cooperator – they personally experience lower $f$ and/or lower $p$ than average. By contrast, a highly desirable individual has a better chance of either finding a partner at all (higher $f$) and/or finding a partner who is a good partner (higher $p$). As such, highly desirable individuals will have less stake in the current partners because they can replace them more easily.

**2d) Summary of model variations**

Once again, these models contain few assumptions and are thus very general and apply to many situations. Further, they do not depend on the partner finding out that they have been saved, and thus work just as well if the saving is anonymous. As such, saving is not "reciprocated" (e.g., "I save you this time, you save me next time"). Instead, an existing reciprocal relationship – or the possibility of one in the future – gives agents a stake in their partner's welfare, such that it is worth saving that partner so they can continue the reciprocal relationship. When agents can replace incapacitated partners, this reduces the stake in one's partners, but considerable stake can remain, especially in stable relationships. Some individuals may have more (or less) stake in their partners than do other individuals, because some individuals find it harder (or easier) to attract good replacements.

## 3) Model A: Helping that affects a partner's fecundity instead of its condition/survival

Suppose that instead of helping a partner to remain in good condition (e.g., alive, healthy, able to reciprocate, available), organisms can pay a cost to increase the partner's fecundity. Is it still worthwhile to help? In this case, it depends on whether the focal individual can expect either itself or its kin to interact with that recipient's offspring (e.g., vertical transmission of a reciprocity-based relationship). I will examine each of these separately.

### 3a) If the focal individual interacts with recipient's offspring

If the relationship is vertically transmitted, then it can be worthwhile to help good partners to reproduce, as long as their offspring are more likely than average to also be good reciprocators. By investing in a good partner's fecundity, an organism makes itself more likely to experience reciprocation in the future because it increases the cooperativeness of its pool of future partners. The better that a partner's offspring are relative to other potential partners, the more worthwhile it is to help that partner breed, because then one's pool of future partners is even better. Conversely, if one's pool of future partners is already very cooperative (e.g., good cooperators are very common, other mechanisms of assortment exist), then there is little benefit in helping a partner breed because their offspring will be no better than any randomly-selected new partner.

The benefits of helping a partner breed depend greatly on many assumptions. These assumptions include the method of assortment with offspring, other mechanisms of assortment available (e.g., partner choice), the resemblance between parent and offspring, the size of one's pool of future partners, the number of future partners (which depends on age of focal individual), the number of current offspring a partner has, litter size, the marginal impact of each new offspring on the probability of assortment with them, how assortment is represented mathematically (e.g., exponential vs. another function), and so on. As such, modeling this is far beyond the scope of the present article. For now, it is enough to know that organisms can have a stake in a good partner's fecundity, as long the focal individual or its kin will then interact with that partner's nicer-than-average offspring.

### 3b) If focal individual does not interact with recipient's offspring

Conversely, if neither the focal individual nor its kin will interact with the recipient's offspring, or if those offspring are no more cooperative than other expected partners, then there are no return benefits from helping a partner reproduce. In such cases, organisms have no stake in their partner's fecundity.

If a partner's reproduction reduces its own ability to invest in reciprocity, then a partner's fecundity actually reduces the focal individual's stake in that partner. If each unit of fecundity reduces a partner's remaining ability to reciprocate by proportion $f$ ($0 \leq k \leq 1$), and the partner has $m$ units of fecundity, then at any given time point the focal individual's stake is only $(1-k)^m$ of what it would have for a nulliparous partner. Thus, the focal individual will only be willing to pay cost $a$ to help keep its partner alive and in good condition when:

$$(1-k)^m(b-c) > a(1-w) \qquad \text{(Inequality S10)}$$

In fact, the focal individual might even prefer that its partner not reproduce at all. Comparing the payoffs with nulliparous partners in Inequality 1 versus with fecund partners in Inequality S10, it

is worthwhile for the focal individual to pay cost $h$ to prevent its partner from reproducing whenever:

$$(b\text{-}c)(1\text{-}(1\text{-}k)^m) > h(1\text{-}w) \qquad \text{(Inequality S11)}$$

Inequalities S10 and S11 are obviously gross simplifications, for example they assume that all units of fecundity arrive at the same time and instantaneously, reduce the partner's ability to reciprocate proportionally instead of additively, and that they do so indefinitely. In any real system, the exact costs and benefits depend on the nature of parental investment, such as the timing of each unit of fecundity, how the pregnancy or incubation period impacts the partner's ability to reciprocate, and for how long those effects last. Analyzing the multitude of possibilities is obviously far beyond the scope of this paper. For now, this grossly simple model suffices to make two simple points: when reproduction reduces a partner's ability to reciprocate (i.e., high $k$), that causes organisms to 1) have less stake in keeping their partners alive and in good condition; and 2) have a greater interest in preventing that partner from reproducing.

### 4) Model B: Discussion of when observability affects Machiavelli vs. AllD

Figure 1 in the main text presented the zones where observed and anonymous cooperation pay off (i.e., TFT wins), where observed cooperation pays off but anonymous cooperation doesn't (i.e., Mach > TFT > AllD and Mach > AllD > TFT), and where no cooperation pays off (i.e., AllD wins). A general conclusion was that observability generally benefits cooperators.

However, there was one curious result: when paired with TFT, anonymity (not observability) seems to help Mach pay better than AllD: as observability ($x$) goes down, Mach pays better than AllD at lower and lower values of $b$ (left side of Fig. S1a). Why? At low values of $b$, cooperation does not pay off. When most rounds are observable, then Mach spends more of its time in unproductive cooperation with TFT, where it would have been better to just defect. However, when most rounds are anonymous it spends more of its time suckering its TFT partner. Mach thus does better than AllD against TFT under low observability, because Mach never gets caught suckering its partner whereas AllD eventually does. This counter-intuitive effect of anonymity only holds for agents paired with a Tit-for-Tat partner. By contrast, when paired with a sneaky Machiavellian partner, observability helps both Mach and TFT pay better than AllD: as observability ($x$) goes down, both Mach and TFT need higher values of $b$ to beat AllD. This is because low observability means that one's Machiavellian partner defects more of the time, such that higher b is necessary to make it worth cooperating with them.

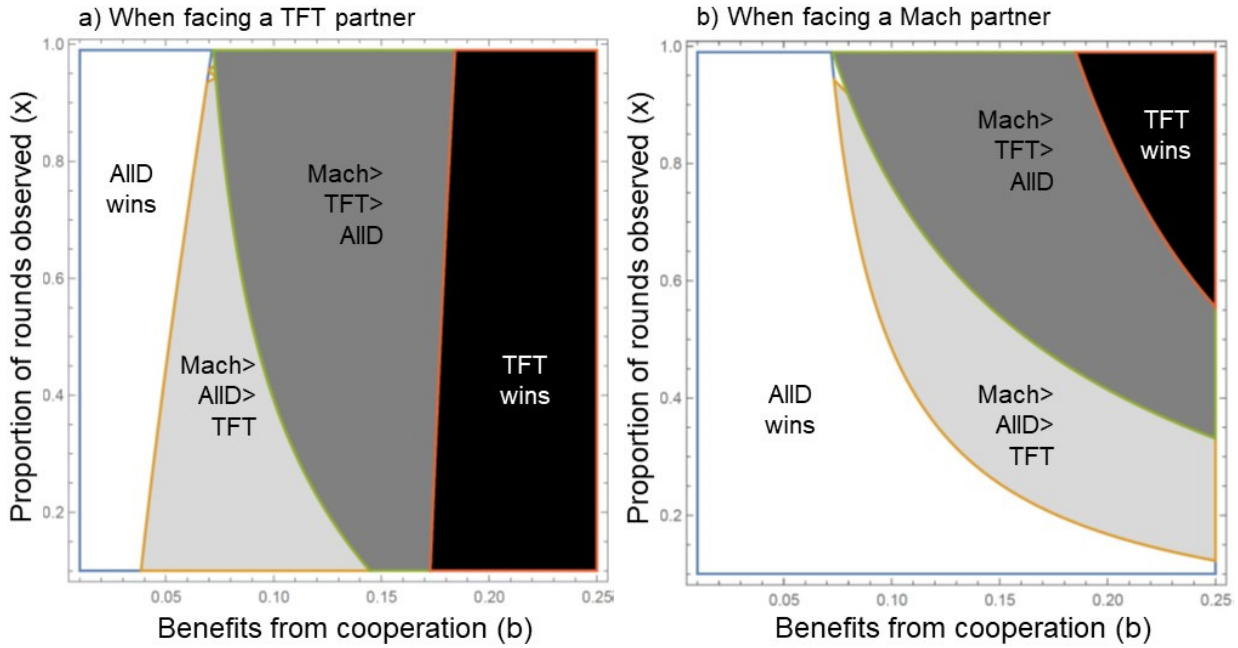


Figure S4: This is Figure 1 from the main text, reprinted here for reference in this discussion so readers don't have to cross-reference back to the main text. Black areas represent conditions where agents have sufficient stake in their partners, such that it pays best to cooperate even when anonymous, for a) Tit-for-tat partners ($T_T>M_T>D_T$); and b) Machiavellian partners ($T_M>M_M>D_M$). Dark gray areas represent conditions where observed cooperation pays off, but anonymous cooperation does not, i.e., a) $M_T>T_T>D_T$; b) $M_M>T_M>D_M$. Light gray areas represent conditions where Mach does best and anonymous cooperation does worst, i.e., a) $M_T>D_T>T_T$; and b) $M_M>D_M>T_M$). White areas represent conditions where no cooperation pays off, i.e., AllD pays best; a) $D_T>M_T>T_T$; and b) $D_M>M_M>T_M$. Parameters displayed are $n$=5, $c$=0.05, $w$=0.75.

## 5) Model B: Varying the model parameters

Figures S5-S10 vary the parameters that were held constant when presenting the findings in the main text, specifically the baseline survivability ($w$), the number of rounds ($n$), and the costs of cooperation ($c$). In each figure, the solid areas represent regions where cooperation pays off with a Tit-for-Tat partner: panels a-c show the conditions where playing Tit-for-Tat pays better than playing AllD, and panels d-f show the conditions where playing Tit-for-Tat also pays better than playing Machiavelli. The latter represent when it pays to help – even when anonymous – to keep one's partner alive.

**5a) Model B: Facing a Tit-for-tat partner, varying everything but the number of rounds (*n*)**

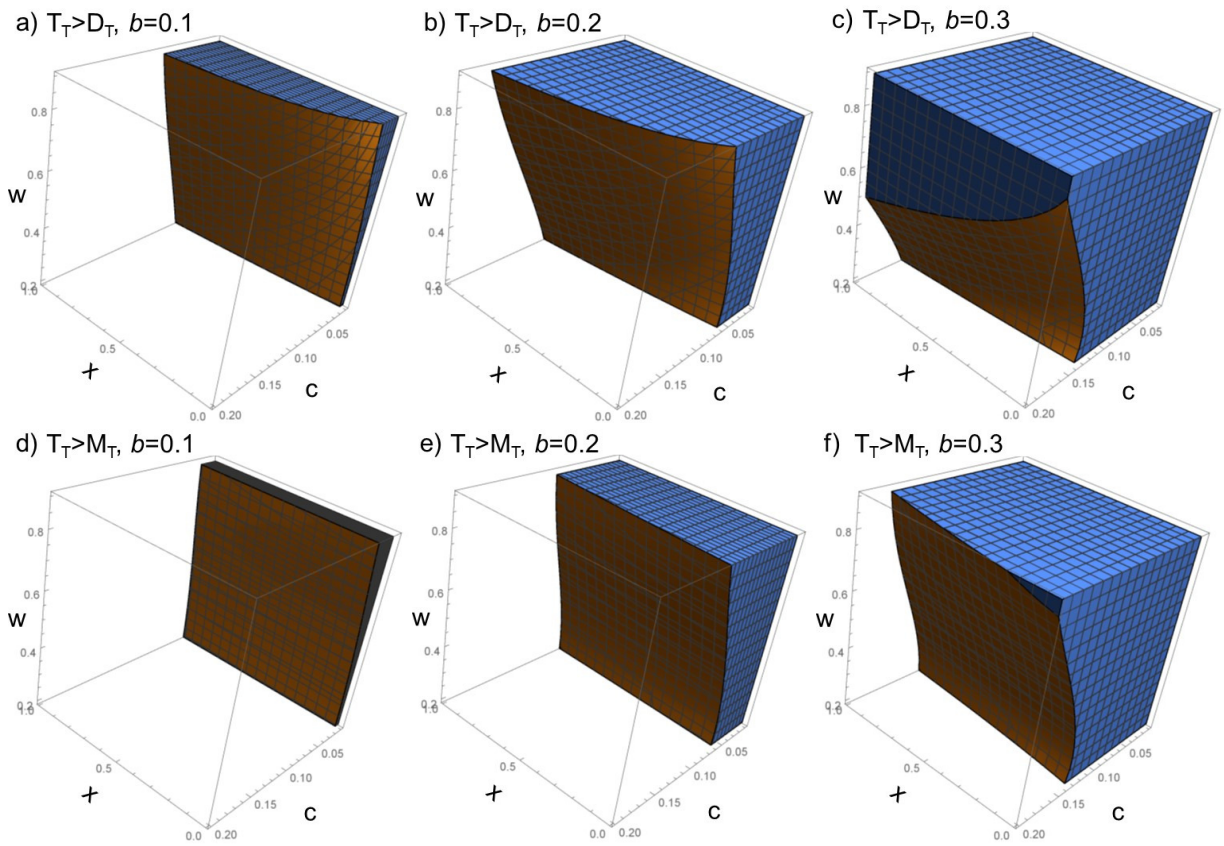


Figure S5: Solid areas represent conditions where it pays to cooperate with a Tit-for-Tat partner. Panels a-c compare when it pays to cooperate in general instead of defect (i.e., $T_T>D_T$); panels d-f examine when it pays to cooperate even when anonymous (i.e., $T_T>M_T$). The axes are the proportion of rounds that are observable (*x*), baseline survival per round (*w*), and the costs of cooperating (*c*); please note that the values of *x* and *c* decrease from left to right. Any region shaded in panels a-c but not in panels d-f is where it pays to cooperate with TFT in observed rounds but not anonymous rounds (i.e., $D_T<T_T<M_T$). Parameters displayed are $b$=0.10 in panels S5a & S5d, $b$=0.20 in panels S5b & S5e, and $b$=0.30 in panels S5c & S5f; $n$=10 in all panels.

Within each panel of Figure S5, cooperation is more likely to pay off as baseline survivability (*w*) increases, and as costs decrease (decreasing *c*). Comparing across panels from a-c and d-f, cooperation pays better as the gains from cooperation (*b*) increase. These patterns are true regardless of whether we compare Tit-for-Tat with AllD (Fig S5, panels a-c) to assess the general benefits of cooperation, or if we compare Tit-for-Tat with Machiavelli (Fig S5, panels d-f) to specifically assess the benefits of cooperating when anonymous (i.e., paying solely to keep one's partner alive).

Increasing observation (*x*) makes playing Tit-for-Tat pay better than All-D (Fig S5 panels a-c), but it does not make playing Tit-for-Tat pay better than Mach (Fig S5 panels d-f). If anything, too much anonymity harms Machiavelli when it plays with a good partner because then Machiavelli defects more often and its partners die.

**5b) Model B: Facing Machiavellian partners, varying everything but number of rounds (*n*)**

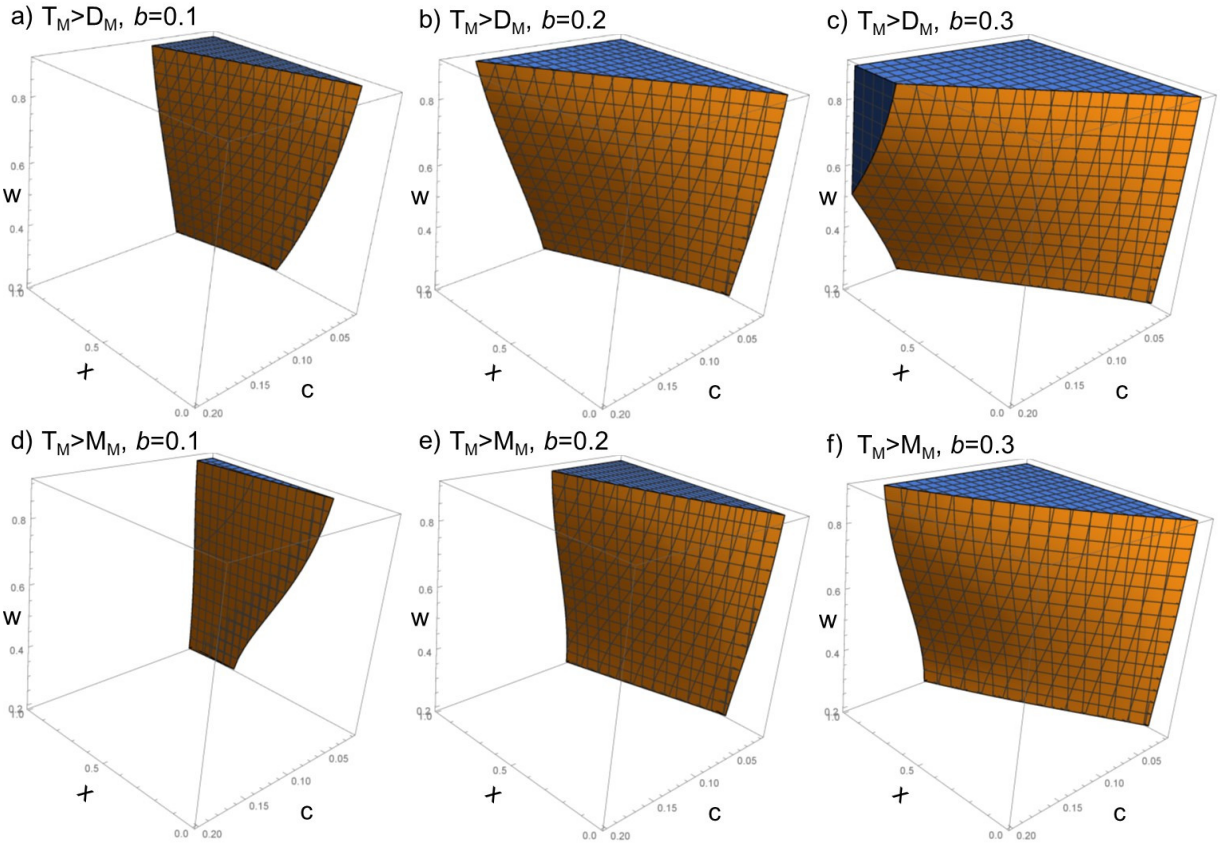


Figure S6: Solid areas represent conditions where it pays to cooperate with a Machiavellian partner. Panels a-c compare when it pays to cooperate in general instead of defect (i.e., $T_M>D_M$); panels d-f examine when it pays to cooperate even when anonymous (i.e., $T_M>M_M$). The axes are the proportion of rounds that are observable ($x$), baseline survival per round ($w$), and the costs of cooperating ($c$); please note that the values of $x$ and $c$ decrease from left to right. Any region shaded in panels a-c but not in panels d-f is where it pays to cooperate with TFT in observed rounds but not anonymous rounds (i.e., $D_M<T_M<M_M$). Parameters displayed are $b$=0.10 in panels S6a & S6d, $b$=0.20 in panels S6b & S6e, and $b$=0.30 in panels S6c & S6f; $n$=10 in all panels.

Within each panel of Figure S6, cooperation is more likely to pay off as baseline survivability ($w$) increases, and as costs decrease (decreasing $c$). Comparing across panels from a-c and d-f, cooperation pays better as the gains from cooperation ($b$) increase. These patterns are true regardless of whether we compare Tit-for-Tat with AllD (Fig S6 panels a-c) to assess the general benefits of cooperation, or if we compare Tit-for-Tat with Machiavelli (Fig S6 panels d-f) to specifically assess the benefits of cooperating when anonymous (i.e., paying solely to keep one's partner alive).

When playing a Machiavellian partner, increasing observation ($x$) makes Tit-for-Tat more likely to pay better than both All-D (Fig S6 panels a-c) *and* Machiavelli (Fig S6 panels d-f). This is different from Figure S3 because a Machiavellian partner spends more of its time cooperating (i.e., is a "good partner" more often) when most rounds are observed, and is thus worth keeping alive only when most rounds are observed.

**5c) Model B: Facing a Tit-for-tat partner, varying everything but costs of cooperation (*c*)**

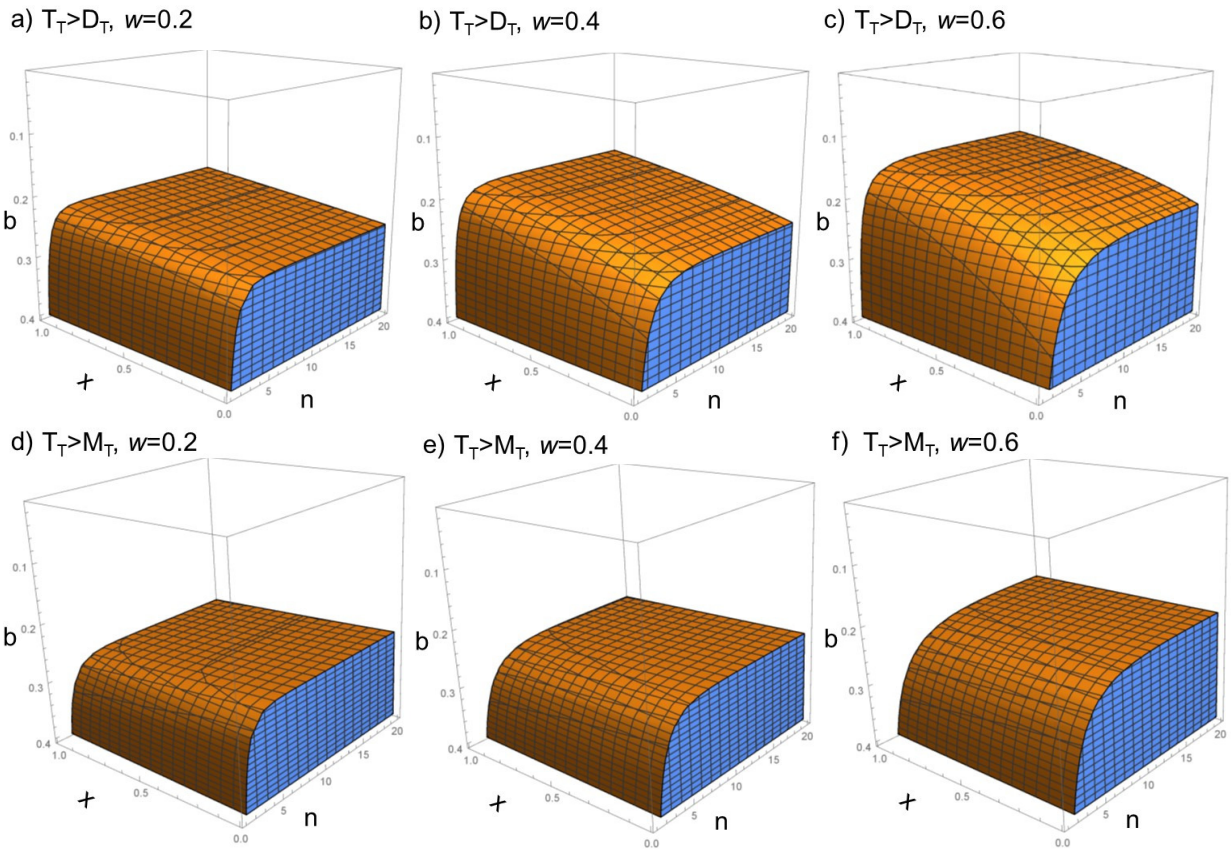


Figure S7: Solid areas represent conditions where it pays to cooperate with a Tit-for-Tat partner. Panels a-c compare when it pays to cooperate in general instead of defect (i.e., $T_T>D_T$); panels d-f examine when it pays to cooperate even when anonymous (i.e., $T_T>M_T$). The axes are the proportion of rounds that are observable (*x*), the number of rounds (*n*), and the gains from cooperating (*b*); please note that the values of *x* decreases from left to right and the values of *b* decrease from top to bottom. Any region shaded in panels a-c but not in panels d-f is where it pays to cooperate with TFT in observed rounds but not anonymous rounds (i.e., $D_T<T_T<M_T$). Parameters displayed are *w*=0.20 in panels S7a & S7d, *w*=0.40 in panels S7b & S7e, and *w*=0.60 in panels S7c & S7f; *c*=0.1 in all panels.

Within each panel of Figure S7, cooperation is more likely to pay off as the number of rounds increase (*n*), and as benefits increase (*b*). Comparing across panels from a-c and d-f, cooperation pays better as baseline survivability (*w*) increases. These patterns are true regardless of whether we compare Tit-for-Tat with AllD (Fig S7, panels a-c) to assess the general benefits of cooperation, or if we compare Tit-for-Tat with Machiavelli (Fig S7, panels d-f) to specifically assess the benefits of cooperating when anonymous (i.e., paying solely to keep one's partner alive).

Increasing observation (*x*) makes playing Tit-for-Tat pay better than All-D (Fig S7 panels a-c), but it does not make playing Tit-for-Tat pay better than Mach (Fig S7 panels d-f). If anything, too much anonymity harms Machiavelli when it plays with a good partner because then Machiavelli defects more often and its partners die.

A close examination of Figure S7 reveals that it does not take many rounds for cooperation to pay off. The variable *n* has its biggest effects on other parameters at small numbers of rounds. For example, if we look at how high *b* has to be for Tit-for-Tat to pay better than Machiavelli (panels S7d-g), we see a big difference between *n*=2 and *n*=4, but little difference between *n*=4 and *n*=6 or even n=20. This means that partners develop a stake in each other's welfare under a wide variety of parameters even if the game will only last a few rounds – it does not take much reciprocity to have a stake in one's partner's welfare. This means that our model is applicable to a wide variety of reciprocal exchange systems, including systems with relatively few instances to help. For example, among the Maasai pastoralists of East Africa, herders form "osotua" relationships with each other whereby they freely give cattle to their partners whenever their partners need cattle (e.g., post-drought) and they themselves have enough to give (Aktipis et al., 2016). These need-based transfers are done without any expectation of debt or repayment – it's enough to know that one's osotua partner would do the same for you when you need cattle. This possibility of reciprocation (i.e., help when you eventually need it) creates a stake in one's partner's welfare, such that the Maasai are willing to give without any debt or repayment. Our model shows that this stake develops even if there are relatively few transfers of cattle – the number of rounds can be very small yet it can still pay to cooperate anonymously, because doing so keeps alive a partner who would help you.

**5d) Model B: Facing a Machiavellian partner, varying everything but the costs of cooperation (*c*)**

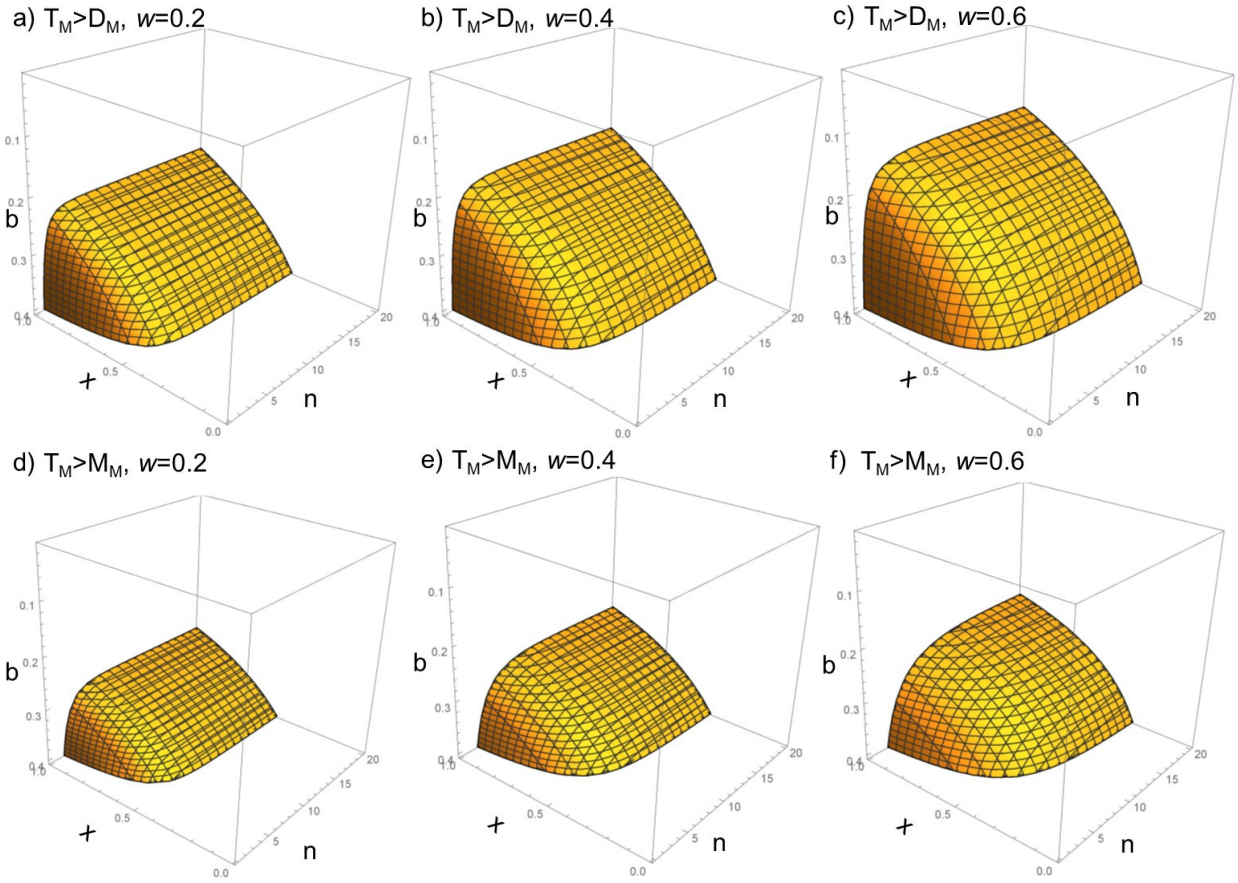


Figure S8: Solid areas represent conditions where it pays to cooperate with a Machiavellian partner. Panels a-c compare when it pays to cooperate in general instead of defect (i.e., $T_M>D_M$); panels d-f examine when it pays to cooperate even when anonymous (i.e., $T_M>M_M$). The axes are the proportion of rounds that are observable (*x*), baseline survival per round (*w*), and the benefits of cooperating (*c*); please note that the values of *x* increases from right to left, and the values of *b* increase from top to bottom. Any region shaded in panels a-c but not in panels d-f is where it pays to cooperate with TFT in observed rounds but not anonymous rounds (i.e., $D_M<T_M<M_M$). Parameters displayed are *w*=0.20 in panels S8a & S8d, *w*=0.40 in panels S8b & S8e, and *w*=0.60 in panels S8c & S8f; *c*=0.10 in all panels.

As in Figure S7, all panels of Figure S8 show that cooperation is more likely to pay off as the number of rounds increase (*n*), and as benefits increase (*b*). Also as in Figure S7, cooperation pays better as baseline survivability (*w*) increases (comparing across panels from a-c and d-f). Furthermore, like Figure S7, most of the gains from increased rounds (*n*) are at lower numbers of rounds. For example, there is a big difference between how easily Tit-for-Tat can win when there are 2 vs. 4 rounds, but little difference between 4 and 6 or even 20 rounds. This shows that it does not take many rounds for organisms to develop a stake in reciprocal partners, even if those partners are sneaky Machiavellians who will defect when anonymous.

Unlike Figure S7, Figure S8 shows that increased observation (*x*) makes playing Tit-for-Tat pay better than All-D (Fig S8 panels a-c) and also makes playing Tit-for-Tat pay better than Mach

(Fig S8 panels d-f). This is because the Machiavellian partners in Figure S8 spend more time cooperating when observation is high, whereas the Tit-for-Tat partners in Figure S5 cooperated regardless of observation. As such, good Tit-for-Tat partners are worth keeping alive regardless of observability or anonymity, whereas sneaky Machiavellian partners are only worth keeping alive when observation is high and they're mostly cooperative.

### 5e) Model B: Anonymous cooperation when *n*=2

Some reciprocal partnerships are short-lived and involve only a few exchanges. Sections 2c and 2d showed that it pays to anonymously help valued partners even if there are relatively few rounds (Figures S5-S6). Here I show that anonymous cooperation can even pay off when the shadow of the future is very short – only two rounds (*n*=2) – provided that the benefits of cooperation (*b*) are sufficiently high, the costs are sufficiently low (*c*), and most rounds are observed (high *x*). Figure S7 presents the parameter regions where cooperation pays off with a Tit-for-Tat partner, and Figure S8 presents the parameter regions where cooperation pays off with a Machiavellian partner.

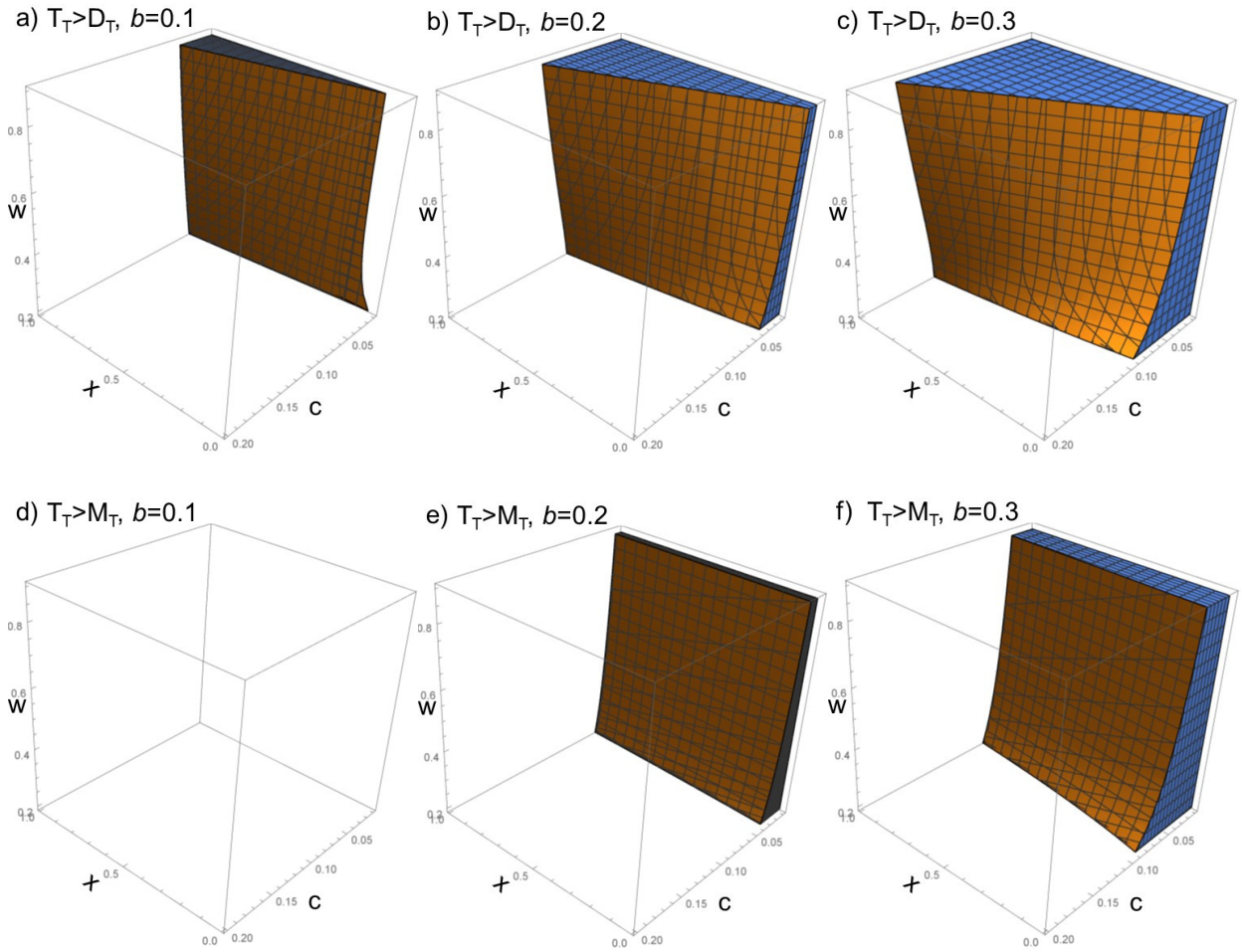


Figure S9: Solid areas represent conditions where it pays to cooperate with a Tit-for-Tat partner when *n*=2. Panels a-c compare when it pays to cooperate in general instead of defect (i.e., $T_T>D_T$); panels d-f examine when it pays to cooperate even when anonymous (i.e., $T_T>M_T$). The axes are the proportion of rounds that are observable (*x*), baseline survival per round (*w*), and the costs of cooperating (*c*); please note that the values of *x* and *c* decrease from left to right. Any region shaded in panels a-c but not in panels d-f is where it pays to cooperate with TFT in observed rounds but not anonymous rounds (i.e., $D_T<T_T<M_T$). Parameters displayed are *b*=0.10 in panels S9a & S9d, *b*=0.20 in panels S9b & S9e, and *b*=0.30 in panels S9c & S9f.

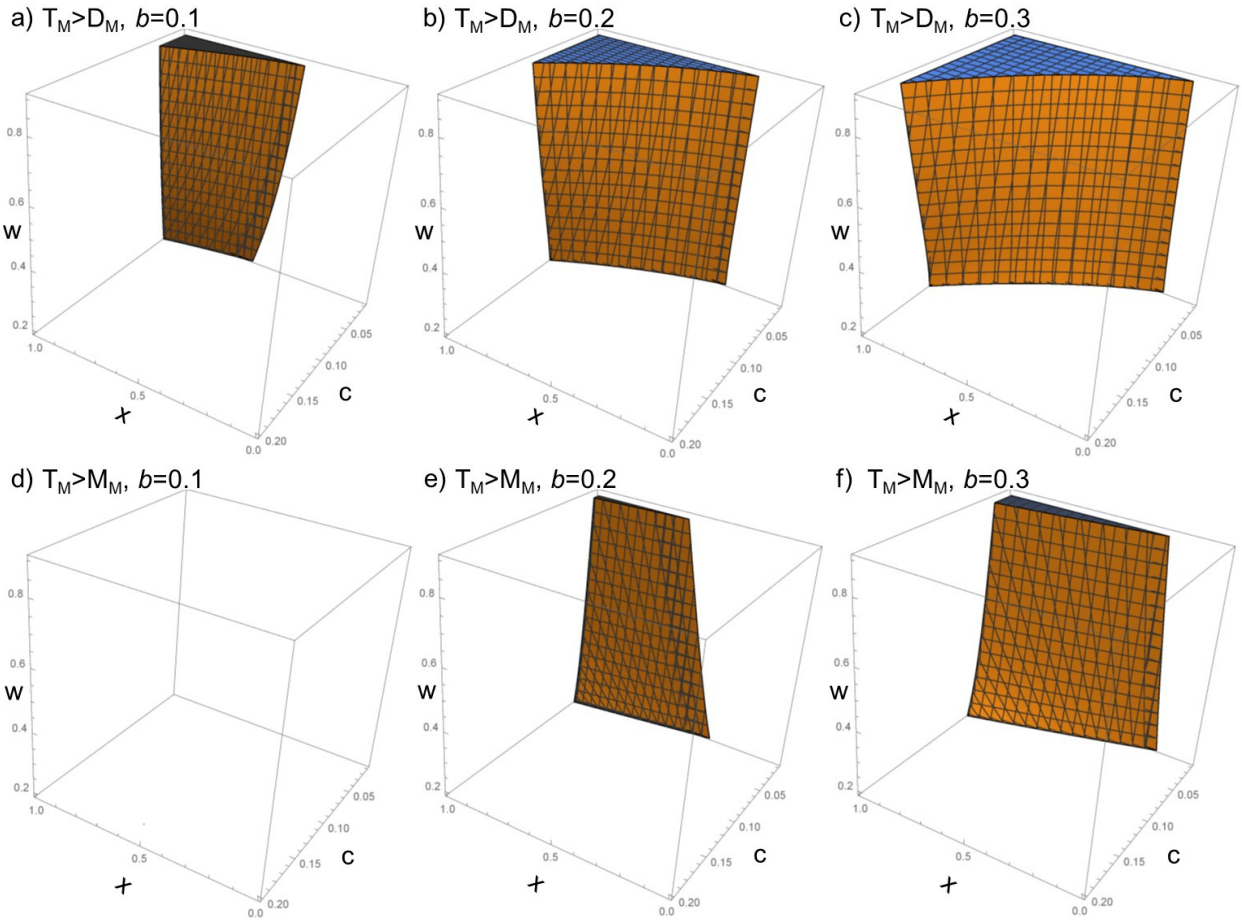


Figure S10: Solid areas represent conditions where it pays to cooperate with a Machiavellian partner when $n$=2. Panels a-c compare when it pays to cooperate in general instead of defect (i.e., $T_M>D_M$); panels d-f examine when it pays to cooperate even when anonymous (i.e., $T_M>M_M$). The axes are the proportion of rounds that are observable ($x$), baseline survival per round ($w$), and the costs of cooperating ($c$); please note that the values of $x$ and $c$ decrease from left to right. Any region shaded in panels a-c but not in panels d-f is where it pays to cooperate with TFT in observed rounds but not anonymous rounds (i.e., $D_M<T_M<M_M$). Parameters displayed are $b$=0.10 in panels S10a & S10d, $b$=0.20 in panels S10b & S10e, and $b$=0.30 in panels S10c & S10f.

Figures S9 and S20 show that cooperation can pay off even in very short-lived partnerships (panels a-c). Organisms can even have a stake in such short-term partners, such that it can pay to help when anonymous, provided that the benefits of cooperation are high enough (compare the zero cooperation in panel d in both Figures with the cooperation in panels e-f).

The results in these short-term partnerships are largely similar to those from longer partnerships (e.g., Figures S5-S8):

- Both observed and anonymous cooperation pay better than defection when benefits are large ($b$) and costs are small ($c$);
- Observability ($x$) generally makes observed and anonymous cooperation pay better than defection. The exception is that low observability particularly hurts Machiavellians paired with Tit-for-Tat, because Machiavelli defects when anonymous and its good partner dies (Figure S8 panels e-f);

- High baseline survival ($w$) generally – but not always – makes cooperation pay better than defection (see below).

The only new finding occurs when agents are faced with a Tit-for-tat partner and most rounds are anonymous (low $x$). In these cases, Tit-for-Tat is more likely to outperform AllD and Machiavelli when baseline survivability $w$ is *low*, not high (Figure S9 panels b-c and e-f). This is because AllD suckers its partner in round 1, and Machiavelli usually does too if most rounds are anonymous. When baseline survivability is high, these "suckered" partners are likely to survive despite being suckered, and will thus survive to help in round 2. As such, it pays to defect if most rounds are anonymous and baseline survivability is high. By contrast, under low baseline survivability (i.e., high mortality), any "suckered" partner is unlikely to survive to help in round 2, so AllD and Machiavelli have no one to help them in round 2 and are likely to die, whereas Tit-for-Tat keeps its partners alive and thus benefits from their cooperation in round 2.

**5f) Model B: Summary of varying the parameters of the model**

These supplementary results support the results in the main text by showing that our central finding is robust to wide variations in the parameters, namely that it often pays to cooperate with good partners even when anonymous, because doing so keeps those partners alive. This anonymous cooperation – like observed cooperation – pays off better when the benefits of cooperation ($b$) are higher, the costs of cooperation ($c$) are lower, and the "shadow of the future" is high because of many rounds ($n$) and high baseline survivability ($w$).

The supplementary results add nuance to the findings about observability. Unsurprisingly, observability ($x$) helps Tit-for-Tat to beat defectors, regardless of one's partner, because then Tit-for-Tat can detect defectors earlier. Observability also helps Tit-for-tat to outcompete Machiavelli against other Machiavellian partners, because high observability makes such partners sufficiently cooperative to make it worth paying anonymously to keep them alive. By contrast, when paired with a truly cooperative partner, observability does not help Tit-for-Tat outcompete Machiavelli, and may even hinder anonymous cooperation (Fig S4 panels d-f). As long as some potential partners are defectors or sneaky Machiavellians, then observability will generally result in higher observed and anonymous cooperation.

The supplementary results also add some nuance to the findings about the length of interactions ($n$). Cooperation – even anonymous cooperation – can pay off even if there are relatively few rounds. Even two rounds can be sufficient for anonymous cooperation to evolve, provided that observability ($x$) and the benefits of cooperation ($b$) are sufficiently high.

## 6) Model B: Critical Thresholds for Invasion

If we apply the principles of Evolutionarily Stable Strategies, Tit-for-Tat can invade Machiavelli whenever Tit-for-Tat has a higher payoff against Machiavelli than Machiavelli does against itself. The main text and Supplementary show that this is often the case (Figures 1b and panels d-f of S5, S7, and S9): when playing a Machiavellian partner, Tit-for-Tat often has a higher payoff than Machiavelli. In other words, the previous figures in the main text show the conditions under which Tit-for-Tat can invade a population of pure Machiavellians.

The main text shows the strictest conditions for invasion: if there are zero Tit-for-Tat players in the population, such that any Tit-for-Tat player is guaranteed to encounter Machiavelli, Figures 1b, S5, S7, and S9 show the conditions where Tit-for-Tat can still invade. However, if there are at least some Tit-for-Tat players in the population, then Tit-for-Tat can outcompete Machiavelli under a broader range of conditions. How many Tit-for-Tat players must there be for anonymous cooperation to pay off? This is the critical threshold: the frequency of Tit-for-Tat players where organisms have enough stake in their partners, such that it starts to pay to cooperate anonymously. Below this threshold, there is insufficient stake, and anonymous cooperation does not pay off (i.e., Machiavelli outcompetes Tit-for-Tat). Above this threshold, organisms have enough stake such that anonymous cooperation does pay off (i.e., Tit-for-Tat outcompetes Machiavelli).

In this section, I present the critical threshold that Tit-for-Tat must surpass to outcompete Machiavelli, i.e., the proportion of the population that must play Tit-for-Tat in order for Tit-for-Tat to pay better than Machiavelli. Whenever this critical threshold is zero, it means that a single Tit-for-Tat player does better than the Machiavellians in the population (as they do in for some values in Figures 1b, S5, S7, and S9). If this critical threshold is 0.10, it means that Tit-for-Tat pays worse than Machiavelli if the former represents less than 10% of the population, but pays better if they are more than 10% of the population. I also vary the proportion of defectors in the population from 0.0 to 0.3, to show what happens when all three strategies are present in the population.

To test this, I create three new variables for the overall payoff of AllD ($w_D$), Tit-for-Tat ($w_T$), and Machiavelli ($w_M$):

$$w_D = p_D(D_D) + p_T(D_T) + (1-p_D-p_T)(D_M) \quad \text{(Equation S12)}$$

$$w_T = p_D(T_D) + p_T(T_T) + (1-p_D-p_T)(T_M) \quad \text{(Equation S13)}$$

$$w_D = p_D(M_D) + p_T(M_T) + (1-p_D-p_T)(M_M) \quad \text{(Equation S14)}$$

Where $p_D$, $p_T$, and $1-p_D-p_T$ are the proportion of the population playing AllD, Tit-for-Tat, and Machiavelli, respectively. The critical threshold for anonymous cooperation is the proportion of Tit-for-Tat ($p_T$) where $w_T>w_M$; once Tit-for-Tat becomes more common than that, it outcompetes Machiavelli. I also calculate where $w_T>w_D$ to ensure that anonymous cooperation can also beat AllD at those proportions (spoiler alert: it does).

The results show that Tit-for-Tat can often pay better than AllD and Machiavelli, even if other Tit-for-Tat players are rare. In other words, organisms often have sufficient stake in their partners such that anonymous cooperation pays better than anonymous defection or outright defection, even when other anonymous cooperators are rare. Figure S11 compares the payoff of Tit-for-Tat against Machiavelli when some fixed proportion of the population plays AllD ($p_D$=0.0 in panels a-c and $p_D$=0.3 in panels d-f), and the rest either play Tit-for-Tat or

Machiavelli. Figure S12 does the same except it compares the payoffs of Tit-for-Tat against AllD. These figures show that as long as there are some partners around who cooperate when observed (i.e., Tit-for-Tat or Machiavelli), and enough of the rounds are observed, then an organism has enough stake in its partner to keep that partner alive, even if it means cooperating anonymously.

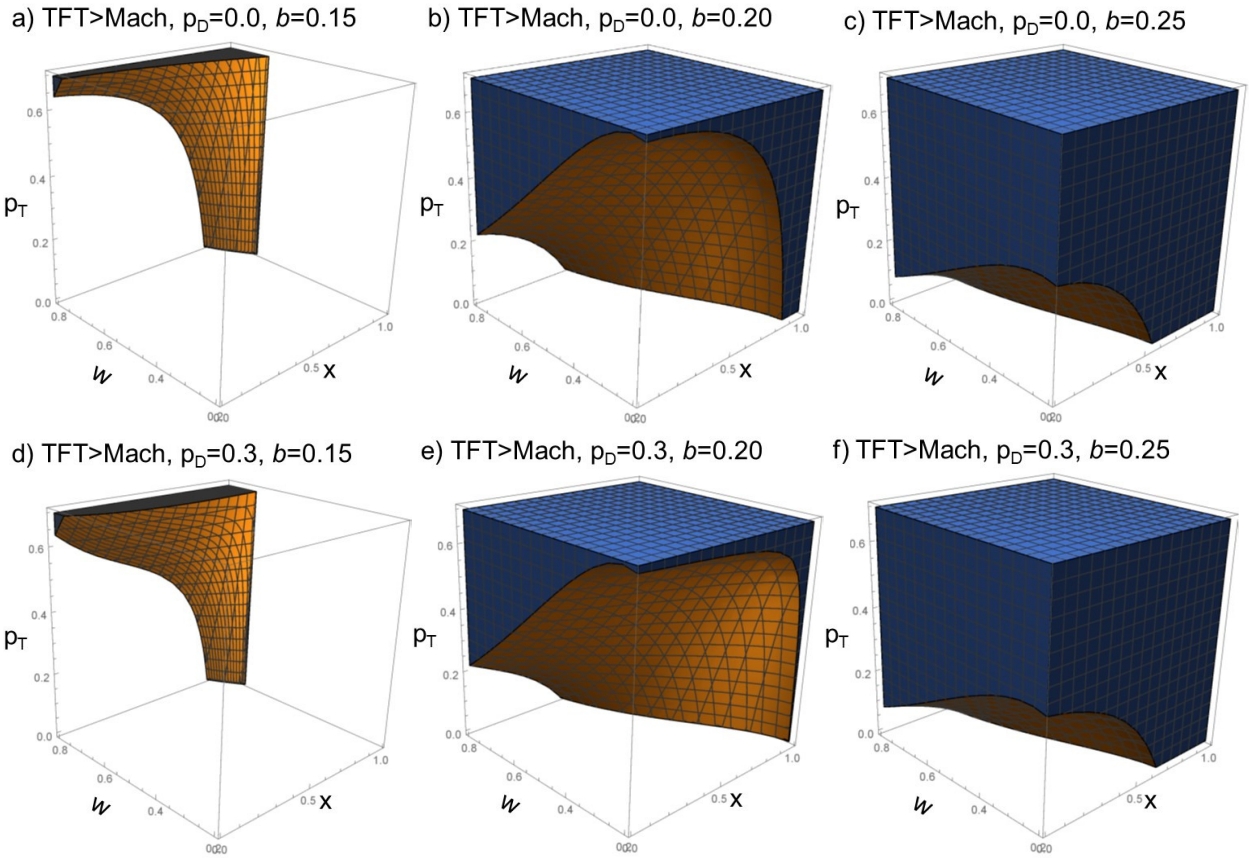


Figure S11: Critical threshold of TFT where TFT starts to pay better than Machiavelli. Solid regions represent parameter combinations where the proportion of TFT ($p_T$) is high enough for TFT to invade Machiavelli; empty regions represent parameter conditions where $p_T$ is too low. Thus, for any given combination of $w$ and $x$, the critical threshold of $p_T$ is represented by the border between the empty and solid regions. The axis extending left is the baseline survival per round ($w$) from 0.2 to 0.8, and the axis extending right is the proportion of rounds that are observable ($x$) from 0.001 to 0.999. Panels a-c are when the proportion of AllD is $p_D$=0.0, panels d-f are when the proportion of AllD is $p_D$=0.3; all others are Machiavelli. Parameters displayed are $b$=0.15 in panels S11a & S11d, $b$=0.20 in panels S11b & S11e, and $b$=0.25 in panels S11c & S11f; $c$=0.05 and $n$=10 in all panels.

As Figure S11 shows, the critical threshold for Tit-for-Tat to invade Machiavelli is lower when baseline survival is higher ($w$), more rounds are observed ($x$), and the benefits for cooperation are higher ($b$). Critical thresholds are also lower when the costs for cooperation ($c$) are lower and the number of rounds ($n$) is higher (results not shown but available upon request).

Surprisingly, the proportion of defectors (relative to Machiavelli) has relatively little impact on the critical threshold (compare panels a-c with d-f). This is because Tit-for-Tat and Machiavelli behave similarly towards AllD: they both start cooperating in observed rounds and defect after

first detecting defection. The main difference is that when observability is low, Tit-for-Tat gets suckered for longer in the anonymous rounds before eventually observing a defection. However, when observability is low, Tit-for-Tat does poorly against Machiavelli as well anyway, so this extra disadvantage has little additional effect.

Figure S12 shows the critical Threshold for Tit-for-Tat to invade AllD. It shows the same patterns as Figure S11 (TFT vs. Machiavelli), except that the critical thresholds are all lower in Figure S12, i.e., it takes fewer Tit-for-Tat players to beat AllD than it does to beat Machiavelli.

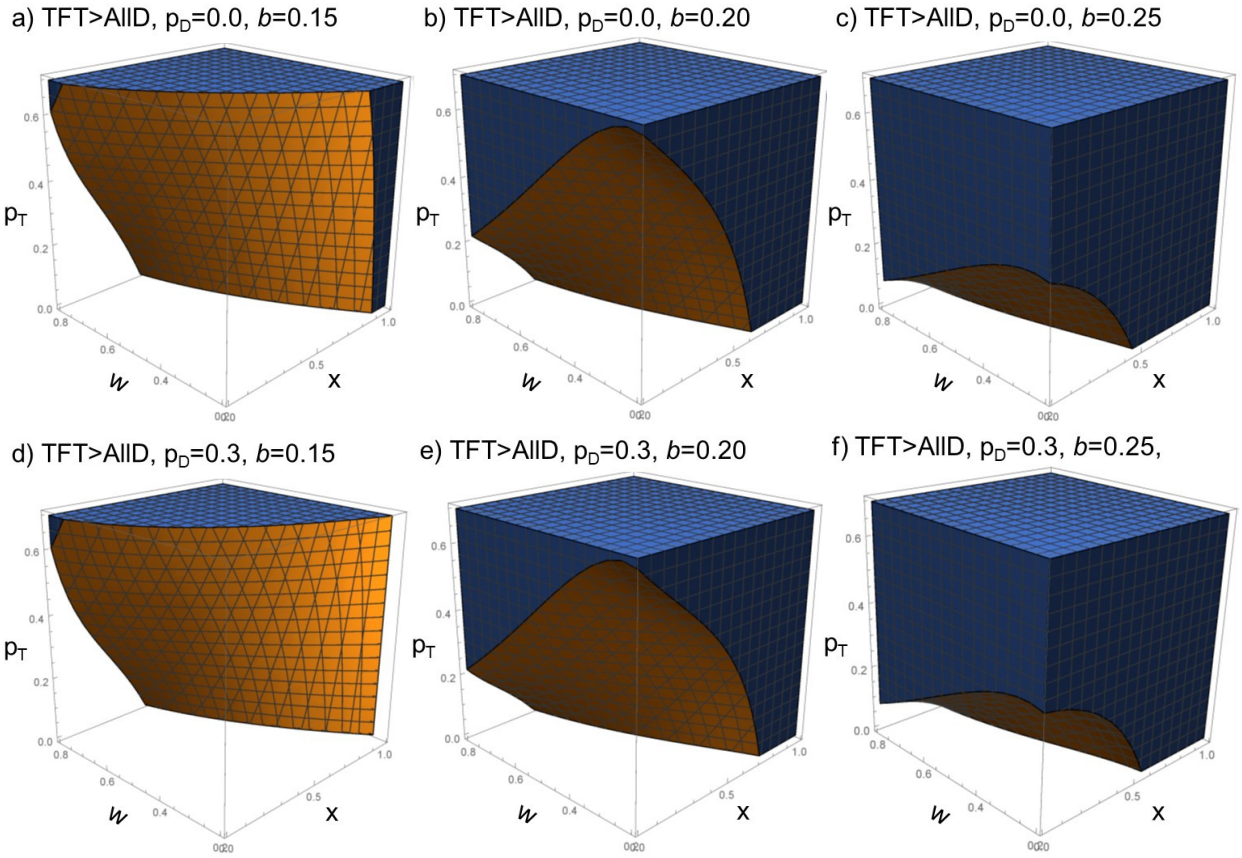


Figure S12: Critical threshold of TFT where TFT starts to pay better than AllD. Solid regions represent parameter combinations where the proportion of TFT ($p_T$) is high enough for TFT to invade AllD; empty regions represent parameter conditions where $p_T$ is too low. Thus, for any given combination of $w$ and $x$, the critical threshold of $p_T$ is represented by the border between the empty and solid regions. The axis extending left is the baseline survival per round ($w$) from 0.2 to 0.8, and the axis extending right is the the proportion of rounds that are observable ($x$) from 0.001 to 0.999. Panels a-c are when the proportion of AllD is $p_D$=0.0, panels d-f are when the proportion of AllD is $p_D$=0.3; all others are Machiavelli. Parameters displayed are $b$=0.15 in panels S12a & S12d, $b$=0.20 in panels S12b & S12e, and $b$=0.25 in panels S12c & S12f; $c$=0.05 and $n$=10 in all panels.

## 7) Model B: Variation of the model – scaling the costs and benefits

In the main text, I assume simple additive costs and benefits of cooperation. This raises the possibility of endpoint effects, if help given or received can bring one's survivability to 0 or 1, respectively. As a robustness check to prevent endpoint effects, here I present a variation on the model where help given and received affects one's *residual* mortality, i.e., whatever mortality remains between 1 and w. In this version, all costs ($c$) and benefits ($b$) are scaled by 1-w, such that survival can never equal one. This scaling affects the costs & benefits in the current round, as well as one's partner's survival in previous rounds. The payoffs are:

$$\mathrm{D_D} = \prod_{t=1}^{n} w \quad \text{Equation S15}$$

$$D_T = \prod_{t=1}^{n}(w + (1-w)b(1-x)^{t-1}(w-(1-w)c)^{t-1}) \quad \text{Equation S16}$$

$$D_M = \prod_{t=1}^{n}(w + (1-w)bx(1-x)^{t-1}w^{t-1}) \quad \text{Equation S17}$$

$$T_D = \prod_{t=1}^{n}(w - (1-w)c(1-x)^{t-1}(w+(1-w)b)^{t-1}) \quad \text{Equation S18}$$

$$T_T = \prod_{t=1}^{n}(w + (1-w)(b-c)(w+(1-w)(b-c))^{t-1}) \quad \text{Equation S19}$$

$$T_M = \prod_{t=1}^{n}(w + (1-w)(bx-c)(x(w+(1-w)(b-c)) + (1-x)(w+(1-w)b))^{t-1}) \quad \text{Equation S20}$$

$$M_D = \prod_{t=1}^{n}(w - (1-w)cx(1-x)^{t-1}w^{t-1}) \quad \text{Equation S21}$$

$$M_T = \prod_{t=1}^{n}(w + (1-w)(b-xc)(x(w+(1-w)(b-c)) + (1-x)(w-(1-w)c))^{t-1}) \quad \text{Equation S22}$$

$$M_M = \prod_{t=1}^{n}(w + (1-w)x(b-c)(x(w+(1-w)(b-c)) + (1-x)w)^{t-1}) \quad \text{Equation S23}$$

This variation produces results almost identical to those in the main text. When paired with a defector, it pays best to defect: $\mathrm{D_D} \geq \mathrm{M_D} \geq \mathrm{T_D}$ for all parameter values, so there is no stake without reciprocity. When paired with a Tit-for-Tat partner, observed and anonymous cooperation can pay off under many circumstances (Figures S13-S15). Both observed and anonymous cooperation pay better when there are high benefits from cooperation ($b$) and low costs of cooperation ($c$) (Figure S13-S14). As before, observability ($x$) helps Tit-for-Tat pay better than AllD, and pay better than Machiavelli when facing another Machiavellian (Figure S12-S14).

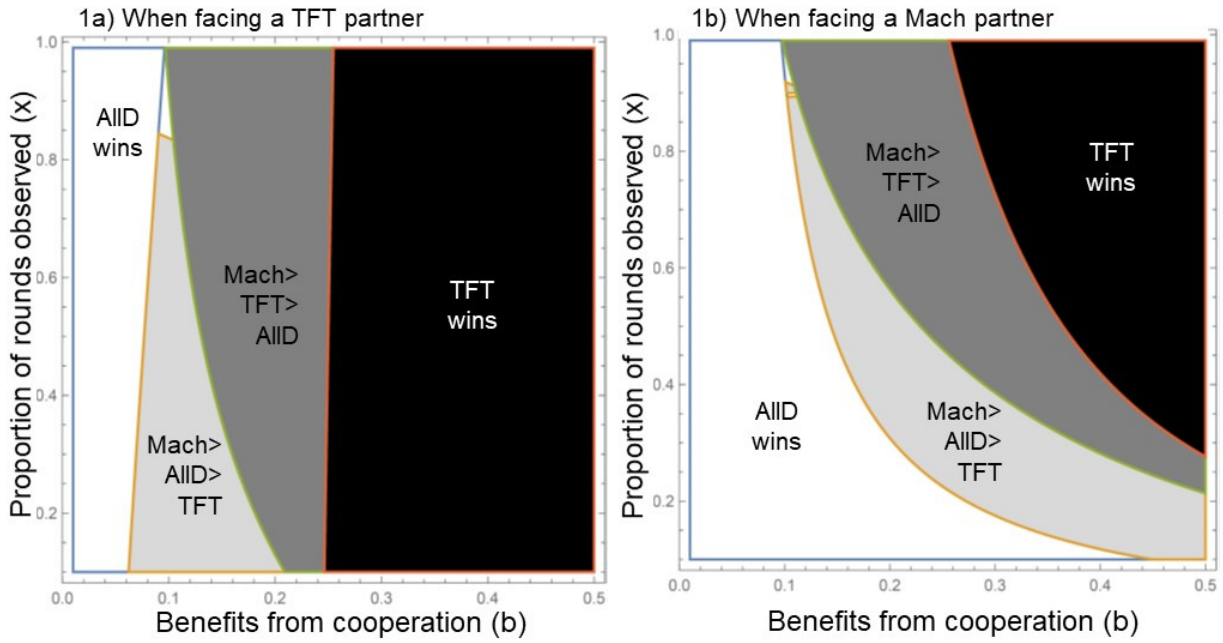


Figure S13: Strategies which perform best in the alternate model with costs and benefits scaled by residual mortality (1-*w*), when paired with a) TFT partners; and b) Machiavellian partners. Black areas represent conditions where agents have sufficient stake in their partners, such that it pays best to cooperate even when anonymous, i.e., a) $T_T>M_T>D_T$ and b) $T_M>M_M>D_M$. Dark gray areas represent conditions where observed cooperation pays off, but anonymous cooperation does not, i.e., a) $M_T>T_T>D_T$; and b) $M_M>T_M>D_M$. Light gray areas represent conditions where Machiavelli does best and anonymous cooperation does worst, i.e., a) $M_T>D_T>T_T$; and b) $M_M>D_M>T_M$). White areas represent conditions where no cooperation pays off, i.e., AllD pays best: a) $D_T>M_T>T_T$; and b) $D_M>M_M>T_M$. Mutual cooperation earns $w+b-c$. Parameters displayed are $n$=5, $c$=0.05, $w$=0.5.

Two results differ slightly in this residual model from the results in the main text. First, the range of parameters for cooperation is slightly more restrictive in this residual model than in the model in main text, hence the difference in scale in S13 and S14. Scaling by residual mortality effectively reduces both the costs and benefits of cooperation, so the gains from cooperation are lower relative to the baseline probability of dying.

Second, in the main text, higher survivability usually makes observed and anonymous cooperation pay better than defection, excepting only some combinations of low survivability and low observability (Figure S8). By contrast, in this model of residual mortality, Tit-for-Tat is more likely to pay better than Machiavelli when baseline survivability ($w$) is low (Figures S14-S15 panels d-f). This is because when baseline survivability ($w$) is high, scaling the costs and benefits by residual mortality (1-$w$) greatly reduces the actual impact of cooperation. When the experienced costs and benefits are lower, anonymous cooperation is less likely to pay off because it's less worth keeping a good partner around.

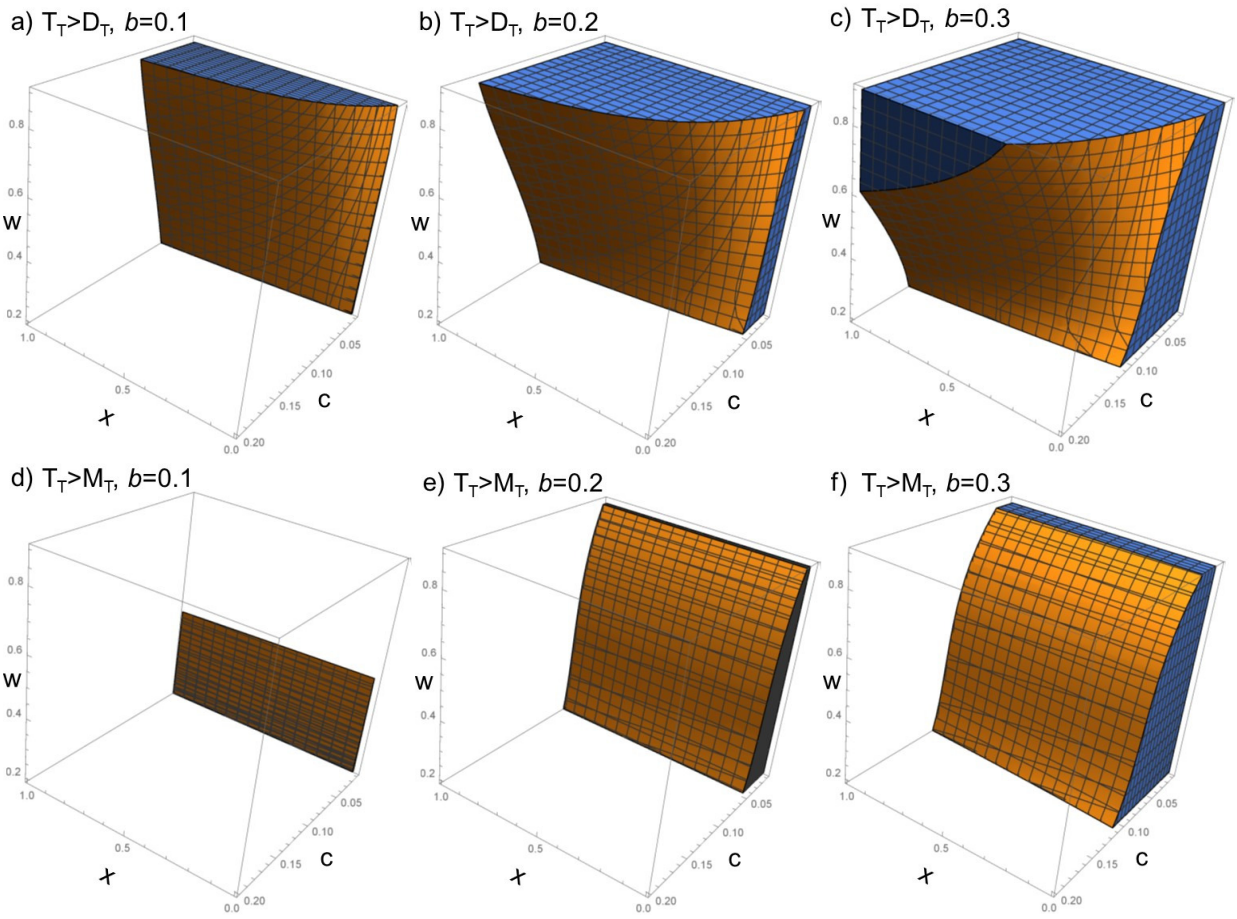


Figure S14: Results of scaled model with Tit-for-Tat partners. Solid areas represent conditions where it pays to cooperate with a Tit-for-Tat partner. Panels a-c compare when it pays to cooperate in general instead of defect (i.e., $T_T$>$D_T$); panels d-f examine when it pays to cooperate even when anonymous (i.e., $T_T$>$M_T$). The axes are the proportion of rounds that are observable ($x$), baseline survival per round ($w$), and the costs of cooperating ($c$); please note that the values of $x$ and $c$ decrease from left to right. Any region shaded in panels a-c but not in panels d-f is where it pays to cooperate with TFT in observed rounds but not anonymous rounds (i.e., $D_T$<$T_T$<$M_T$). Parameters displayed are $b$=0.10 in panels S14a & S14d, $b$=0.20 in panels S14b & S14e, and $b$=0.30 in panels S14c & S14f; $n$=10 in all panels.

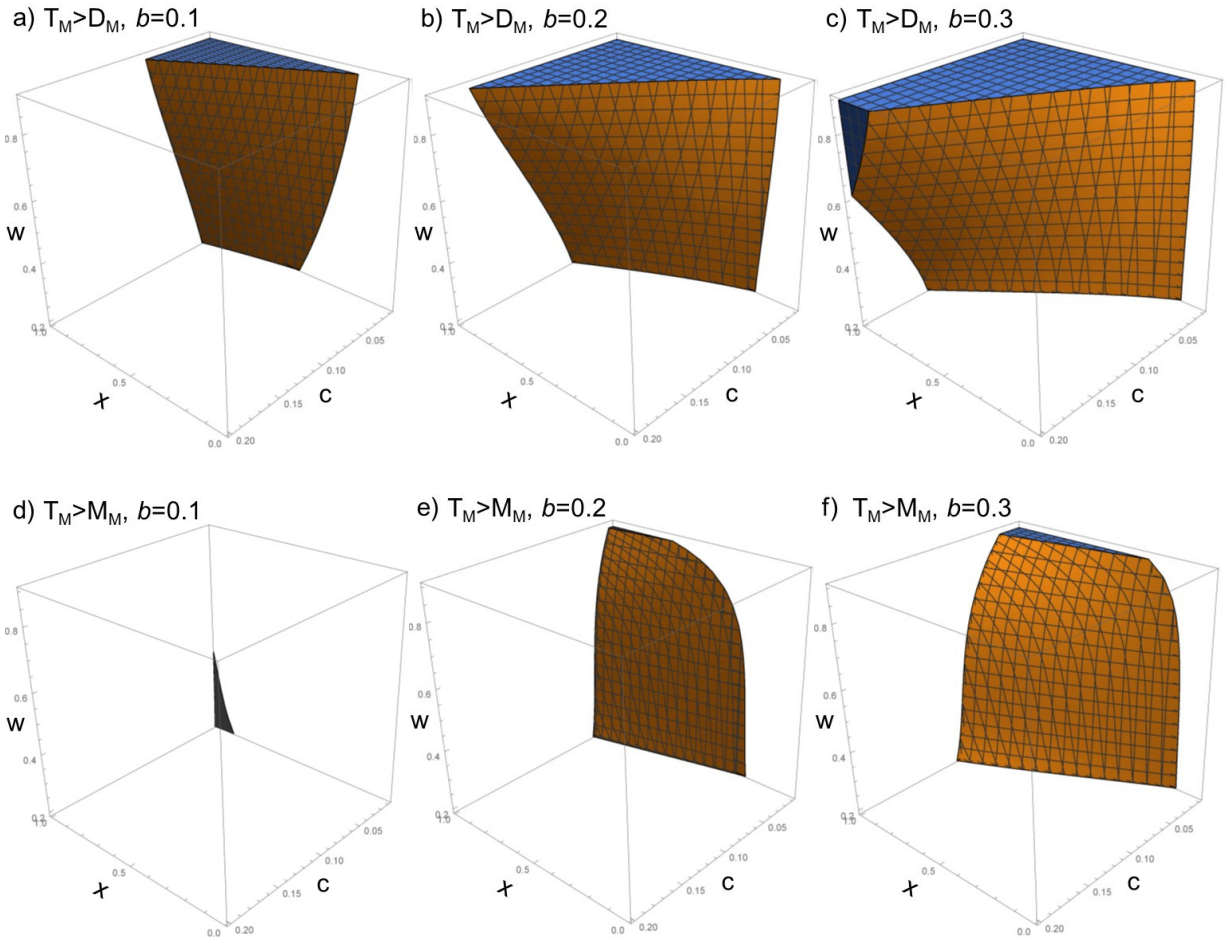


Figure S15: Results of scaled model with Machiavellian partners. Solid areas represent conditions where it pays to cooperate with a Machiavellian partner. Panels a-c compare when it pays to cooperate in general instead of defect (i.e., $T_T>D_T$); panels d-f examine when it pays to cooperate even when anonymous (i.e., $T_T>M_T$). The axes are the proportion of rounds that are observable ($x$), baseline survival per round ($w$), and the costs of cooperating ($c$); please note that the values of $x$ and $c$ decrease from left to right. Any region shaded in panels a-c but not in panels d-f is where it pays to cooperate with TFT in observed rounds but not anonymous rounds (i.e., $D_T<T_T<M_T$). Parameters displayed are $b$=0.10 in panels S15a & S15d, $b$=0.20 in panels S15b & S15e, and $b$=0.30 in panels S15c & S15f; $n$=10 in all panels.

*Summary of Scaled Model*

This scaled model eliminates the zeroes and ones that could occur with the model from the main text, yet still produces the same general results (albeit under fewer conditions). Thus, the principle behind the model is robust with respect to this assumption of unscaled benefits. Even when costs and benefits are scaled according to the residual mortality, it still often pays to cooperate anonymously with a good partner (Figure S13a, Figure S14 panels d-f). It can even pay to cooperate anonymously with a Machiavellian partner (Figure S13b, Figure S15 panels d-f), provided the cost is not too high and enough rounds are observed for that partner to be good most of the time.