

Provided for non-commercial research and educational use only.  
Not for reproduction or distribution or commercial use.



This article was originally published in a journal published by Elsevier, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues that you know, and providing a copy to your institution's administrator.

All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at:

<http://www.elsevier.com/locate/permissionusematerial>

# Perfect reciprocity is the only evolutionarily stable strategy in the continuous iterated prisoner's dilemma

Jean-Baptiste André<sup>a,c,\*</sup>, Troy Day<sup>b</sup>

<sup>a</sup>*Department of Biology, Queen's University, Kingston, Ont., Canada*

<sup>b</sup>*Department of Mathematics and Statistics, Queen's University, Kingston, Ont., Canada*

<sup>c</sup>*Instituto Gulbenkian de Ciência, Oeiras, Portugal*

Received 18 April 2006; received in revised form 12 February 2007; accepted 13 February 2007

Available online 3 March 2007

## Abstract

Theoretical studies have shown that cooperation tends to evolve when interacting individuals have positively correlated phenotypes. In the present article, we explore the situation where this correlation results from information exchange between social partners, and behavioral flexibility. We consider the game 'continuous iterated prisoner's dilemma'. The level of cooperation expressed by individuals in this game, together with their ability to respond to one another, both evolve as two aspects of their behavioral strategy. The conditions for a strategy to be evolutionarily stable in this game are degenerate, and earlier works were thus unable to find a single ESS. However, a detailed invasion analysis, together with the study of evolution in finite populations, reveals that natural selection favors strategies whereby individuals respond to their opponent's actions in a perfectly mirrored (i.e., correlated) fashion. As a corollary, the overall payoff of social interactions (i.e., the amount of cooperation) is maximized because couples of correlated partners effectively become the units of selection.

© 2007 Elsevier Ltd. All rights reserved.

**Keywords:** Game theory; Continuous prisoner's dilemma; Reciprocity; Iterated game; Negotiation

## 1. Introduction

A common metaphor for studying the evolution of cooperation is the Prisoner's dilemma, a game with two strategies: cooperate or defect. In this game, defection yields a larger payoff than cooperation, and should hence be favored by natural selection. When this game is played between two individuals repeatedly, however (the Iterated Prisoner's Dilemma; IPD), a strategy called tit-for-tat is a potential outcome of evolution (Trivers, 1971; Axelrod, 1984; Nowak and Sigmund, 1992). Individuals using this strategy start by cooperating and then express, in each round, the action expressed by their partner in the previous round.

In nature, social interactions are rarely all or nothing, and the IPD metaphor has been extended into a continuous version (Wahl and Nowak 1999a,b; Killingback and Doebeli, 2002). In each round, individuals offer a quantitative amount of resources to their partner. This comes at a cost to them but benefits the recipient. Their partner then responds with a counter offer, and the game is iterated back-and-forth for some, potentially indefinite, period of time. Each player accrues a payoff during each round of the game, as a function of the two players' actions in that round.

It has been difficult to obtain a comprehensive picture of evolution in the continuous IPD because many of the available results stem from computer simulations. A recent review (Doebeli and Hauert, 2005) notes that the results of these previous analyses are complex and varied, but they nevertheless were able to identify some conditions under which cooperation appears to evolve. A complete picture is still lacking, however, because an evolutionarily stable strategy has not yet been identified for this game. Here, we

\*Corresponding author. Laboratoire Éco-Anthropologie et Ethnobiologie - UMR CNRS 5145 - Musée de l'Homme. 7 place du Trocadéro 75116 Paris, France. Tel.: +351 21 8861 698.

E-mail addresses: [jeanbaptisteandre@gmail.com](mailto:jeanbaptisteandre@gmail.com) (J.-B. André), [tday@mast.queensu.ca](mailto:tday@mast.queensu.ca) (T. Day).

focus on the special case where (i) in choosing their move for a given round, individuals only make use of the previous round as a source of information and (ii) the number of rounds of interaction between individuals is large and leads to a stable agreement. We demonstrate that, in these circumstances, the only ESS of the continuous IPD corresponds to perfect quantitative reciprocity.

**2. Methods**

Let us consider a simple social interaction between two players. Player 1 donates an amount of energy  $u_1$  to player 2, which generates a benefit  $B(u_1)$  but comes at a cost  $C(u_1)$ , and vice versa. The payoff for player 1 is  $w_1 = B(u_2) - C(u_1)$ , and that for player 2 is  $w_2 = B(u_1) - C(u_2)$ . This

simple game has been previously referred to as the continuous prisoner’s dilemma (Wahl and Nowak 1999a,b; Killingback and Doebeli, 2002). In the absence of correlation between players, the evolutionarily stable strategy is to donate nothing (i.e., cooperation is absent), because donation only comes at a cost to the individual expressing it.

Each individual is also characterized by a ‘response rule’, giving its donation as a function of the donation made by its partner. When two players interact, they repeatedly modulate their respective donation according to their opponent’s past donation, and may ultimately reach a stable agreement point (Fig. 1a). If the number of interactions between the two individuals is sufficiently large, the pre-equilibrium interactions can be neglected, and the average payoffs obtained by both players are

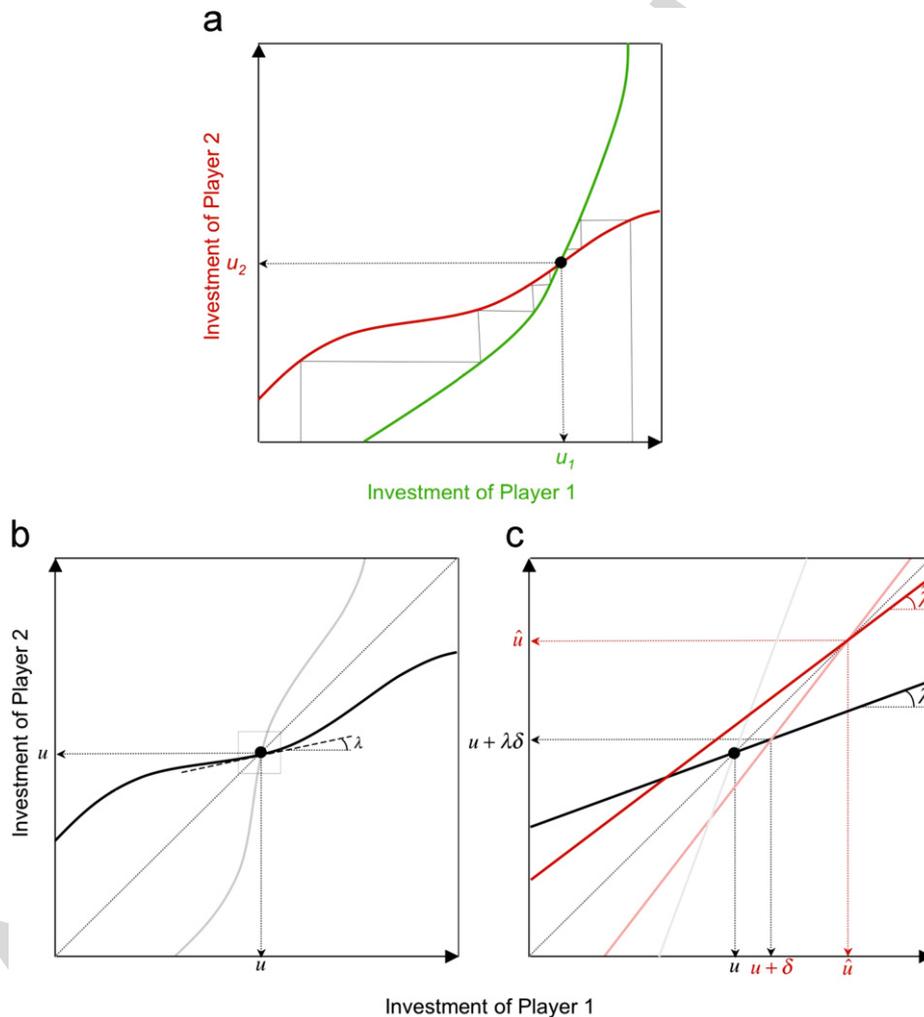


Fig. 1. Iteration and the finding of a stable agreement. (a) Successive bouts of interaction between two different players (green and red). Here the green player makes a first move, which is assessed by the red player and responded to by a ‘countermove’, and so on until a stable agreement point is reached (black circle). (b) Successive bouts of interaction between two individuals playing the same response rule. The two representations of the rule, corresponding to each individual, are symmetric of one another along the diagonal. The stable agreement point is on the diagonal, and can be reached only if the response rule intersects the diagonal with a slope lower than one ( $\lambda < 1$ ). The investment made by each player at this agreement is  $u$ . A necessary relationship must hold between  $\lambda$  and  $u$  for the response rule to be evolutionary stable (Eq. (2)). (c) Detail of (b) around the agreement point, showing the effect of a mutant response rule (dark and light red). The resident rule has slope  $\lambda$  and the mutant  $\hat{\lambda}$ . When a mutant plays against a resident, the mutant ends up donating  $u + \delta$  at the stable agreement and the resident  $u + \lambda\delta$ ; when two mutants encounter they both end up donating  $\hat{u}$ .

approximately  $\bar{w}_1 = B(u_2) - C(u_1)$  and  $\bar{w}_2 = B(u_1) - C(u_2)$ , where the  $u_i$  correspond to the stable agreement donations of each player. An alternate, mathematically equivalent, possibility is to assume that players first undertake a negotiation phase where information is exchanged, and truly play the game only once they have reached a stable agreement (McNamara et al., 1999; Taylor and Day, 2004). Because the negotiation framework raises the complex question of contracts, in the remainder of the paper we focus on the iteration interpretation. We refer to this paradigm of social interaction as the continuous iterated prisoner's dilemma (CIPD). Our aim is to determine how the response rule is expected to evolve, and thereby to predict whether or not cooperation is expected to result.

### 3. Results

#### 3.1. Degeneracy of the ESS condition

Suppose a population consists of a resident type and a rare mutant type at frequency  $q$ . When two resident individuals encounter, they ultimately reach a stable agreement where both give the same donation,  $u$ , and receive  $P_{rr} = B(u) - C(u)$ . The slope  $\lambda$  of their response rule around this agreement is called their behavioral 'responsiveness'. In other words, the resident response rule, as complex as it may be, is defined by two parameters only: its slope  $\lambda$  and donation level  $u$  at the intersection with the diagonal (see Fig. 1b). The response rule of mutant individuals is defined by two parameters as well. However, rather than being defined in absolute terms, it is in part defined relative to the resident (see Fig. 1c). First, mutant responsiveness around the diagonal is defined in absolute terms as  $\hat{\lambda}$ . Second, we suppose that mutant donation, at the stable agreement with a resident, is increased by a small amount  $\delta$  relative to the usual resident donation  $u$ .

In mixed encounters, when mutant's deviation,  $\delta$ , is small, the resident ends up donating approximately  $u + \lambda\delta$  at the stable agreement (Fig. 1c). Therefore, the mutant receives  $P_{mr} = B(u + \lambda\delta) - C(u + \delta)$ , and the resident  $P_{rm} = B(u + \delta) - C(u + \lambda\delta)$ . In mutant–mutant interactions, the donation made by each mutant at stable agreement is less straightforward to derive. However, from the definition of resident and mutant response rules, simple algebra shows that a pair of mutants necessarily donate  $\hat{u} = u + \delta \cdot (1 - \lambda\hat{\lambda}) / (1 - \hat{\lambda})$  at their stable agreement (Fig. 1c). The payoff obtained by mutants in mutant–mutant interactions is thus  $P_{mm} = B(\hat{u}) - C(\hat{u})$ . Therefore, on average, resident individuals receive a payoff  $P_r = (1 - q) \cdot P_{rr} + q \cdot P_{rm}$ , mutants receive  $P_m = (1 - q) \cdot P_{mr} + q \cdot P_{mm}$ , and, to first order in  $\delta$ , the difference between the two is given by

$$P_m - P_r = \left[ (1 - q)(\lambda b - c) + q(\hat{\lambda} b - c) \frac{1 - \lambda}{1 - \hat{\lambda}} \right] \delta + O(\delta^2), \quad (1)$$

where  $c \equiv dC/du$  and  $b \equiv dB/du$  are the marginal cost and benefit arising from an increase in one's partner's donation, and one's own donation, respectively, evaluated around the stable agreement reached by residents ( $u$ ). We assume that the second derivatives of  $B$  and  $C$  around this agreement are such that the second order terms,  $O(\delta^2)$ , are negative. This analysis is valid provided  $\hat{\lambda} \neq 1$ , because  $\hat{u}$  is not defined otherwise (note that  $\hat{\lambda}$  need not be close to  $\lambda$ ). Furthermore, the above analysis assumes that when two identical individuals encounter, they end up reaching a stable agreement where they both play the same strategy (where the response rule intersects the diagonal; Fig. 1b). More generally, however, this is not the only possibility. There can be *more than one agreement on the diagonal*, only one of which is actually reached depending on initial moves, or there can be *no agreement on the diagonal*. The above analysis can be extended to both cases and the results are unaffected. However, the analysis cannot be extended to the third case where there is *no stable agreement at all*. Our model is therefore restricted to response rules that lead to a stable agreement, and do not generate a perpetual instability of players' donations.

According to the usual definition of an evolutionarily stable strategy (ESS), the resident is a local ESS iff for any weak-effect mutant there exists a frequency  $q$  below which the mutant is counter-selected (Maynard Smith and Price, 1973). From Eq. (1), and as long as  $\hat{\lambda} \neq 1$ , this definition leads to the simple ESS condition:

$$\lambda b - c = 0. \quad (2)$$

Condition 2 is analogous to Hamilton's rule (Hamilton, 1964), except that correlation between partners is due to plasticity instead of genetic relatedness. Donations come at a direct cost to individuals but generate an indirect benefit via partner responsiveness: each unit of energy  $\varepsilon$  donated to partner triggers the donation of  $\lambda\varepsilon$  in return. As a result, donation is large at ESS when responsiveness  $\lambda$  is itself large.

However, in contrast with genetic relatedness, the phenotypic correlation between partners is not an external parameter, but is rather one aspect of the social behavior itself. In this respect, condition 2 is degenerate as there are typically a continuum of values of  $\lambda$  and  $u$  that satisfy it: a behavioral strategy can yield any donation at stable agreement, and yet be an ESS, as long as a corresponding responsiveness is exhibited as well. This degeneracy has been previously observed in continuous games (Taylor and Day, 2004), and explains why earlier works on the CIPD were unable to derive a single ESS (Wahl and Nowak, 1999a,b; Killingback and Doebeli, 2002). However, this finding does not imply that all evolutionarily stable strategies are equally likely endpoints of evolution. Next we present three lines of evidence demonstrating that, out of all possible combinations of  $\lambda$  and  $u$  satisfying Eq. (2), it is perfect reciprocity (i.e.,  $\lambda = 1$ ) that ultimately prevails.

### 3.2. The effect of mutant–mutant interactions

First of all, the payoff obtained by mutants when playing against each other can strongly affect their success in a resident population (see Axelrod, 1984; Nowak and Sigmund, 1992; Nowak et al., 2004). A thorough analysis must take into account the full expression of the difference between mutant and resident payoffs (Eq. (1)), instead of its limit when mutants become virtually absent. This is especially important when the resident is an ESS because, by definition of an ESS, the disadvantage of mutants in encounters with residents is then a second-order term only (Eq. (2)). Performing this analysis shows that all but one ESS combinations of  $\lambda$  and  $u$  are susceptible to the invasion by some weak-effect mutants (i.e., mutants with small  $\delta$ ). The ESS condition ( $\lambda b - c = 0$ ) guarantees that the resident is the best strategy against itself. Therefore, for any mutant, there exists a frequency below which this mutant is disfavored (Maynard Smith and Price, 1973). Yet, in practice, most of the strategies matching this condition are susceptible to mutant invasion for the opposite reason: for any arbitrarily small frequency, there exist some mutants that receive a net advantage at this frequency. Mathematically, this result stems from the expression for the difference between mutant and resident payoffs (Eq. (1)). At any ESS defined by Eq. (2), this difference is given by  $P_m - P_r = q(\lambda b - c)(1 - \lambda)/(1 - \hat{\lambda})\delta + O(\delta^2)$ . Therefore, for any arbitrarily small mutant frequency  $q$ , there exist some mutants with a sufficiently small deviation  $\delta$  such that  $P_m - P_r > 0$  (i.e., the mutant is favored by selection). The only exception, and thus the only strategy that can never be invaded by weak-effect mutants is the strategy called ‘perfect reciprocity’ with  $\lambda = 1$ . This particular ESS leads to a stable agreement  $u^*$  satisfying  $b - c = 0$  (Eq. (2)), and therefore it maximizes the overall payoff of the interaction.

### 3.3. Perfect reciprocity in finite populations

A second line of evidence stems from an analysis of evolution in finite populations paralleling previous work on the iterated prisoner’s dilemma (Nowak et al., 2004). In finite populations, an ESS can be characterized as a strategy that, once fixed, results in all other strategies having a fixation probability that is less than that of a neutral allele (Rousset and Billiard, 2000; Proulx and Day, 2001; Nowak et al., 2004). For populations of any finite size,  $N$ , the only strategy that satisfies this stochastic ESS condition in the CIPD is again perfect reciprocity (see Appendix A1). Thus, among the infinite number of strategies satisfying Eq. (2), all strategies except perfect reciprocity are susceptible to the appearance of mutant alleles that have a selective advantage.

### 3.4. Perfect reciprocity in simulations

The final line of evidence comes from simulations (see Appendix A2). We simulated the evolution of linear

response rules, the possible slopes of these rules varying between  $-2$  and  $+2$ . When the product of the slopes of each partner’s response rule is either larger than 1 or lower than  $-1$ , the intersection of the two rules is an unstable agreement. In this case, the stable agreement of the interaction is reached at the boundary of the space of possible agreements (see Appendix A2 for details). In the simulations, like in the mathematical analysis, we could not consider situations where no stable agreement at all can be reached.

In simple simulations, we considered the successive replacement of one strategy by another, according to a mathematical criterion based on Wahl and Nowak (1999a) (see Appendix A2). Fig. 2 plots two typical evolutionary trajectories of  $\lambda$  in these simulations, starting from  $\lambda = -1$  and ultimately reaching  $\lambda = 1$ . Once the strategy corresponding to perfect reciprocity appears in the population, it quickly spreads to fixation.

In more complex simulations, we considered the presence of polymorphism in the population, in a stochastic individual-based model. We started with perfect reciprocity as a fixed resident. Just like TFT in the prisoner’s dilemma, perfect reciprocity is a neutral-ESS: any mutant response rule that yields the same stable agreement as perfect reciprocity attains the same payoff (see also Taylor and Day, 2004). Therefore, at equilibrium, a range of mutually neutral strategies are present in the population (results not shown). However, these simulations reveal that perfect reciprocity is always stably maintained at an intermediate frequency in the long run, because the emergence of neutral strategies opens an ecological niche for defectors, which in turn generate a selective pressure stabilizing perfect reciprocity.

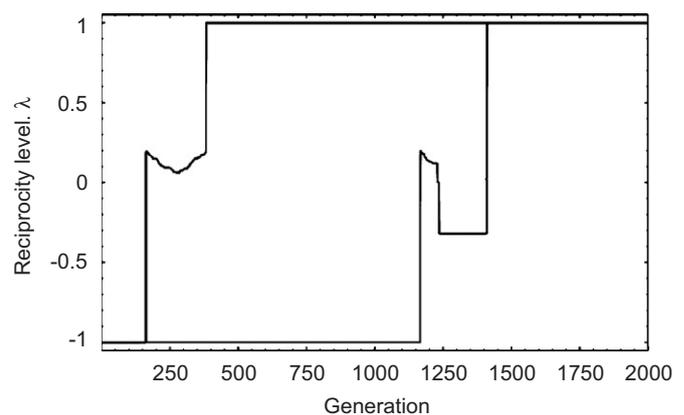


Fig. 2. Evolution of responsiveness of linear rules. Two independent evolutionary paths are presented. Responsiveness evolves via stochastic replacement of residents by mutants (see Appendix A2). Intercept,  $\rho$ , of response rules also evolves (not shown). In each turn of the simulation, a mutation is introduced. In 89% of the cases, the mutation increases  $\lambda$  or  $\rho$  by a small amount drawn from a uniform distribution between  $-0.01$  and  $0.01$ ; in 10% the mutant has a totally random phenotype, and in 1% of the cases the mutant’s strategy is global reciprocation. The benefit and cost functions are  $B(u) = 2u$  and  $C(u) = u^2$ , respectively.

### 3.5. The ‘mirror rule’

The above analyses reveal that a stable endpoint of evolution in the CIPD is necessarily a strategy whereby individuals mirror each other. This condition is local in that it characterizes players’ behavior when donations are close to the stable agreement point, but not more globally (see Appendix A5). However, the simplest way for a player to meet this local reciprocity condition is to be globally reciprocating, i.e., always respond to any donation by the exactly same donation in return, which parallels in a quantitative way the evolution of reciprocity in discrete games (Trivers, 1971; Axelrod, 1984; Nowak and Sigmund, 1992). The drawback of such global reciprocation is that it does not allow players to approach the optimal donation ( $u^*$ ) when starting elsewhere. Hence, the evolution of global reciprocity also requires the evolution of opting for the optimal donation on the first move (Wahl and Nowak, 1999a,b). We then call this strategy the ‘mirror rule’. Consider the mirror rule at frequency  $f$ , in competition with an alternative strategy  $y$ . When  $y$  interacts with a mirror rule,  $y$  ends up playing  $\hat{u}$ , as does the mirror rule (by definition). When a  $y$  individual plays against another  $y$ , both players end up playing  $\hat{u}$  as well. Therefore, the average payoff for  $y$  is  $\hat{P} = B(\hat{u}) - C(\hat{u})$ , and the average payoff of the mirror rule is  $P = f \cdot [B(u^*) - C(u^*)] + (1 - f) \cdot [B(\hat{u}) - C(\hat{u})]$ . Since  $u^*$  maximizes  $B(u) - C(u)$ ,  $P$  is always larger than  $\hat{P}$ . As a result, when starting from any initial frequency,  $f$ , the mirror rule rises to fixation. Therefore, individuals playing this rule not only resist invasion by any weak-effect mutant, but they resist invasion by all mutants, regardless of their starting frequency. Note that, as a result, when the mirror rule is allowed to appear then no other strategy is evolutionarily stable according to the usual ESS definition.

### 3.6. Reciprocity via ‘adjustment rules’

We also considered the more general case where individuals are able to choose their next donation as a function of both their opponent’s and their own past donation (Nowak and Sigmund, 1993), by examining response rules of the form  $u'_s = f_s(u_s, u_p)$ , where  $u_s$  and  $u_p$  are the donations made by a focal individual and her partner. The same results hold in this case as well: full behavioral correlation emerges between social partners and causes social interactions to be optimized (see Appendix A3). Interestingly, a continuum of response rules result in the display of such correlated behavior, the simplest of which consists in donating the optimum amount  $u^*$  on the first move, and then adjusting subsequent moves as a function of the divergence with partner, i.e., following:

$$u'_s = u_s + \gamma(u_p - u_s), \quad (3)$$

where partner-responsiveness  $\gamma$  can take any strictly positive value between zero and one. These rules range from responding exclusively to partner’s previous donation

( $\gamma = 1$ , which corresponds to the mirror rule), to responding almost exclusively to one’s own previous donation ( $\gamma \rightarrow 0$ ). Interestingly, none of these rules correspond to the special case where individuals respond only to their past payoff (Killingback and Doebeli, 2002). In fact, payoff-responding strategies are always susceptible to invasion by at least some weak-effect mutants (see Appendix A3).

We call ‘adjustment rules’ the family of responsive behaviors defined by Eq. (3). Individuals playing these rules do not respond to each move of their opponent by the exact same move, and therefore adjustment rules do not cause mirroring behaviors per se. However, in all cases, individuals playing adjustment rules eventually end up making the same donation as their partner. The mirroring aspect of their behavior thus concerns the stable agreements they reach, and not their proximate behavioral responses.

Although any adjustment rule can resist invasion by mutants of any frequency, natural selection is likely to distinguish among them once error is introduced (see Appendix A.5). With high partner-responsiveness  $\gamma$ , players essentially choose their next donation by imitating their partner’s past move (their behavior is close to the mirror rule). They are thus at risk of responding strongly to either a temporary misinterpretation of their own—the blurred mind, or a temporary mistake of their partner’s—the trembling hand. In contrast, with low responsiveness, players adjust gradually their donation, and converge slowly toward partner’s donation. Therein, they respond to consistent trends in partner’s behaviors, smoothing down temporary noise. Therefore, just like they cause a Pavlovian strategy to be favored in the discrete IPD (Nowak and Sigmund, 1993), communication errors tend to cause natural selection to favor inertia (i.e., low partner-responsiveness  $\gamma$ ) in the CIPD. An information-oriented interpretation of this result is as follows. Information on self is generally more reliable than information on partner because it avoids environmental noise. But knowing oneself per se is not useful in social interactions. Self-knowledge is relevant because it constitutes an indirect, and yet trustworthy, indication of what has likely been happening earlier in the interaction. In other words it allows integrating past information about partner without paying the cost of large memory. Resetting one’s choice in each round (i.e., employing a response rule with  $\gamma = 1$ ) is unwise because it entails the discarding of this information.

## 4. Discussion

In a simple model of social interaction, the continuous prisoner’s dilemma, when the phenotypic correlation between social partners is due to information exchange and behavioral flexibility, any cooperation level can be evolutionarily stable as long as a corresponding type of phenotypic plasticity is expressed as well (Eq. (2)). Partners can ignore each other and fail to cooperate, partially

respond to each other and partially cooperate, or they can perfectly mimic each other and fully cooperate. This can be interpreted in the framework of kin selection theory. The amount of phenotypic correlation between social partners determines the evolutionarily stable level of cooperation (Hamilton, 1964; Taylor and Frank, 1996). When both the level of cooperation and the amount of phenotypic correlation between partners can evolve, there are then an infinite number of evolutionarily stable strategies, each characterized by a given phenotypic correlation and the corresponding level of cooperation (see also Taylor and Day, 2004). However, a careful analysis reveals that behavioral strategies that cause individuals to perfectly match their opponents (i.e., that generate a perfect phenotypic correlation) are, in fact, the only strategies able to resist any mutant invasion. Any other strategy is inevitably at risk of invasion by leaving open the possibility for some mutants to obtain a particularly large payoff against each other. In brief, we have shown that a continuous equivalent of the tit-for-tat is likely to evolve in the CIPD.

The mechanism explaining this result has a parallel in the discrete IPD. Although unconditional defection is an ESS in the discrete IPD, tit-for-tat is nevertheless able to invade if it is present at high enough frequency because tit-for-tat individuals obtain a very large payoff against each other (Axelrod, 1984; Nowak and Sigmund, 1992; Nowak et al., 2004). However, in the continuous version of the game, the potential appearance of small effect mutants makes this mechanism even stronger. By definition, at any ESS, the disadvantage of mutants when playing against residents is a second-order term of mutant deviation. Therefore, for weak-effect mutants, the difference of payoffs in the rare encounters against mutants becomes the leading term determining the fate of mutants (see Eq. (1)). As a result, most evolutionarily stable strategies are, in fact, susceptible to mutant invasion, even in infinite populations, as long as mutant frequency is not strictly equal to zero.

In the discrete IPD a further analysis has shown that a strategy whereby individuals respond both to their own and their partner's past donation could also be an outcome of evolution (Nowak and Sigmund, 1993). We considered this more general case in the CIPD as well (see also Killingback and Doebeli, 2002). We showed that evolution should yield the fixation of any strategy whereby players adjust their behavior, from one round to the next, by incrementing donation as a function of the current divergence with partner (Eq. (3)). Furthermore, among these strategies, communication errors tend to favor those that consist in adjusting donation with inertia, because inertia allows responding only to consistent trends in partner's behaviors, and not to temporary noise.

Killingback and Doebeli (2002) suggest that selection could favor strategies whereby players respond to their past payoff, but our analysis suggests that such strategies cannot, in fact, be evolutionarily stable. Payoff-responding strategies are always susceptible to invasion by at least some weak effect mutants (see Appendix A3).

In relation with this negative result about the evolutionary importance of payoff-responding strategies, we want to stress the fact that these strategies should not be considered as continuous equivalent of Pavlovian strategies (Nowak and Sigmund, 1993). The fact that Pavlovian strategies respond to payoff rather than to partner's last move is not the key part of their definition. In fact, a closer examination of the *win-stay/lose-shift* strategy, which was particularly successful in Nowak and Sigmund (1993), shows that it does in fact respond only to partner's last move (*Stay* when partner cooperates, *Shift* when partner defects; see Nowak and Sigmund, 1993). The originality of Pavlovian strategies rather lies in their ability to respond by a change (or not) compared to last move, instead of settling on a new behavior de novo in each round. Payoff-responding strategies respond to payoff, but they do not achieve the essential aspect of Pavlovian strategies: they decide on their investment de novo in each round.

Our model makes two important assumptions that require discussion. First, it considers only situations where both players actually reach a stable agreement (whether this agreement lies at the boundary of the space of possible agreements or not). This first assumption is in fact double. It implies (i) that partners interact for long enough to reach an agreement and (ii) that there actually is such a stable agreement. We believe the second aspect of the assumption to be relatively mild. In most social/biological exchanges, partners eventually reach a stable type of relationship after some bouts of interaction, provided that duration is sufficient. However, one must keep in mind that our model does not offer any prediction in the relatively rare cases of highly versatile interactions. We believe the first bit of the assumption to be more serious. Indeed, our model does not embrace the numerous cases of relatively brief interactions, where partners do not have enough time to reach a stable agreement.

Second, our model assumes that players only make use of the previous round as a source of information to decide on their next donation. As a result, an individual's response to a given move is the same whenever this move takes place. Another way to put it is to remark that the model assumes a perfect auto-correlation between the way an individual responds at one point of the interaction, and the way he responds at any other point. Relaxing this hypothesis would certainly make cooperation less feasible, if not unachievable. Boyd and Lorberbaum (1987) actually showed that the absence of correlation between individuals' strategy at one round and strategy at other rounds renders TFT unstable in the discrete prisoner's dilemma, and we believe that the same holds in the continuous prisoner's dilemma. However, reciprocity has a meaning provided that the way an individual responds to his partner at one point contains some information about the way he will respond in the future, and therefore the existence of a temporal auto-correlation of individuals' response is a biological prerequisite for reciprocity.

In conclusion, the present paper shows that, in the continuous iterated prisoner's dilemma where partners respond to last move and ultimately reach a stable agreement,

natural selection causes individuals to end up mirroring their opponents. However, the proximate behavioral rule employed by individuals does not necessarily consist in mirroring each and every one of their partner's moves. Individuals may instead adjust gradually their behavior, and eventually converge, at stable agreement only, toward the same donation as partner. In all cases, however, because individuals end up matching their opponents, couples of interacting partners effectively become the units of selection, and cooperation is thus maximized by natural selection.

### Acknowledgments

We thank P. Taylor, G. Wild, D. Nagy and A. Gardner for discussions on related issues. We are grateful to C. Hauert for valuable comments on a previous version of this manuscript and to two anonymous reviewers for their constructive remarks.

## Appendix A

### A.1. Evolutionarily stable strategy in finite populations

Consider the Moran model for evolution in a population of size  $N$ . Suppose the population is fixed for the resident strategy, and introduce a single individual mutant. The probability that the mutant ultimately replaces the resident is

$$\rho = \frac{1}{1 + \sum_{k=1}^{N-1} \prod_{i=1}^k g_i/f_i}, \quad (\text{A.1})$$

where  $g_i$  and  $f_i$  are, respectively, the resident's and mutant's fitness when the population contains  $i$  mutants (and  $N - i$  residents) (Nowak et al., 2004). These are equal to:

$$g_i = 1 - w + w \frac{P_{rm}i + P_{rr}(N - i - 1)}{N - 1},$$

$$f_i = 1 - w + w \frac{P_{mm}(i - 1) + P_{mr}(N - i)}{N - 1}, \quad (\text{A.2})$$

where  $w \in [0, 1]$  specifies the contribution of the game to fitness, and the payoffs  $P_{rr}$ ,  $P_{rm}$ ,  $P_{mr}$  and  $P_{mm}$  are given in the methods of the paper. Assuming that the deviation,  $\delta$ , is small, the probability of fixation of the mutant is

$$\rho = \frac{1}{N} - \frac{\delta B}{6N} S_N(u, \lambda, \hat{\lambda}) + O(\delta^2), \quad (\text{A.3})$$

where  $B = w/[1 - w(1 - P_{rr})]$  depends on the basic payoff  $P_{rr}$  obtained by residents against each other, and the function  $S_N(u, \lambda, \hat{\lambda})$  is given by

$$S_N(u, \lambda, \hat{\lambda}) = (b - c) \times \frac{\lambda - \hat{\lambda}}{1 - \hat{\lambda}} \times (N + 1) + 3b \left( \frac{1 - \lambda \hat{\lambda}}{1 - \hat{\lambda}} - \lambda N \right) - 3c \left( \frac{1 - \lambda \hat{\lambda}}{1 - \hat{\lambda}} - N \right), \quad (\text{A.4})$$

where  $b$  and  $c$  are defined as in the main text. This result does not require that the contribution of the game to fitness,  $w$ , be small.

In a finite population of size  $N$ , the resident is a local ESS iff the fixation probability of any weak-effect mutant (Eq. (A3)) is strictly lower than the fixation probability of a neutral allele (Rousset and Billiard, 2000; Proulx and Day, 2001; Rousset, 2003; Nowak et al., 2004; Wild and Taylor, 2004). The quantity  $B$  is non-zero, and therefore if  $\rho < 1/N$  is to hold for all mutant deviations, we require that  $S_N(u, \lambda, \hat{\lambda}) = 0, \forall \hat{\lambda}$ . This requires that  $b = c$  and that  $\lambda = 1$  around the stable agreement.

Note that, as with the deterministic model, evolutionarily stable strategies are “neutral ESSs” in the sense that any alternative strategy yielding the same stable agreement is neutral in a population fixed for the ESS ( $\rho = 1/N$ ).

### A.2. Simulation methods

In simulations, we considered partner-responsive linear response rules, defined by a slope  $\lambda$  and an intercept  $\rho$ . Both the slope and intercept range from  $-2$  to  $2$ . Possible donations made by individuals are restricted to a given range between  $0$  and  $2$ . Stable agreements can be of different types in this case (Fig. A1).

When the product of the slopes of each partner's response rule is between  $-1$  and  $1$  then the stable agreement is found as the actual intersection of the two rules (Fig. A1a).

When the product of the two slopes is larger than  $1$ , the intersection of the two rules is an unstable agreement. The stable agreement of the interaction is then reached at the boundary of the space of possible agreements (Fig. A1b). In this case, there are two possible agreements and the simulation method allows each one of them to be reached with a probability proportional to its basin of attraction (as if the first move was a random variable uniformly distributed between  $0$  and  $2$ ).

When the product of the two slopes is lower than  $-1$ , the interaction has no stable agreement at all (Fig. A1c). This case cannot be considered in our model. When two individuals are in this situation, we assume that they cannot interact socially and thus obtain a nil payoff.

Mutations can be of two types, each occurring with given probabilities. (1) Small effect on  $\rho$  or  $\lambda$ : intercept or slope of mutant follows a uniform distribution around the intercept or slope of the wild-type. (2) Random mutant independent of ancestor: mutant's intercept and slope are chosen randomly on a uniform distribution in the full range of possibilities.

When two individuals interact, their payoff is calculated as the payoff obtained at stable agreement. The benefit and cost functions are  $B(u) = 2u$  and  $C(u) = u^2$ , respectively. The stable agreement of a given interaction is either that given by the intersection of their two response rules, or else it lies on the border of the square of possible donations.

We performed two types of stochastic simulations.

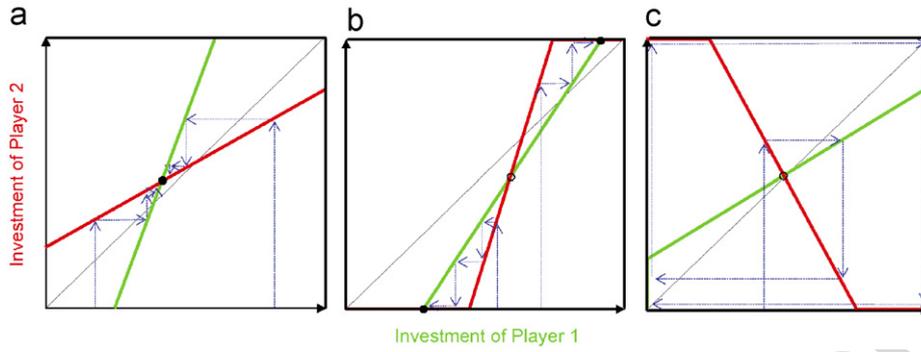


Fig. A1. Examples of encounters between two linear response rules. (a) The product of the two slopes is between  $-1$  and  $1$ . The stable agreement (black dot) is the actual intersection between the response rules. (b) The product of the two slopes is larger than  $1$ . The intersection is an unstable agreement (empty dot); the stable agreements are at the border of the space of possible investments (black dots). (c) The product of the two slopes is lower than  $-1$ . There are no stable agreements in this case, and investment cycles indefinitely.

(1) *Instantaneous replacement*: This simulation method is adapted from Wahl and Nowak (1999a). Polymorphism is neglected, and any mutant that arises either instantaneously replaces the resident or else is eliminated. In each turn, a mutant is introduced (according to the mutation procedure above) and competed against the resident. Three criteria are employed to determine whether the mutant replaces the resident or is eliminated: (i) If  $P_{mr} > P_{rr}$  and  $P_{mm} > P_{rm}$ , then the mutant replaces the resident. (ii) If  $P_{mr} > P_{rr}$  and  $P_{mm} < P_{rm}$ , then the mutant replaces the resident with probability  $p = (P_{mr} - P_{rr}) / (P_{rm} + P_{mr} - P_{mm} - P_{rr})$ . (iii) If  $P_{mr} < P_{rr}$  and  $P_{mm} > P_{rm}$ , then the number of mutants is drawn from a Poisson distribution of mean 1 (in practice the situation where 0 mutants are present is not considered) and their frequency computed by dividing by a parameter representing population size. If the average payoff of mutants at this frequency is larger than that of residents, then the mutant replaces the resident. This last criterion is the only one that differs from Wahl and Nowak (1999a)'s method.

(2) *Individual-based simulations*: Here, polymorphism is allowed. The population has a constant size  $N$ , and generations are non-overlapping. Each individual undergoes mutation followed by reproduction. The fecundity of each individual is equal to a basic fecundity, added to the payoff obtained by the individual when interacting with another random individual drawn from the population. The next generation is then formed by random sampling of  $N$  adults from an effectively infinite pool of propagules.

### A.3. Bivariate response rules

Here, we consider the case where players choose their next donation as a function of both their partner's and their own past donation. The general form of such response rules is  $u'_s = f(u_s, u_p)$ , where  $u_s$  and  $u_p$  are the donations made by a focal individual and her partner. Consider a resident strategy and a rare mutant employing, respectively, the response rules  $f$  and  $\hat{f}$ . Suppose that, when two residents encounter, they end up reaching a stable

agreement where both play  $u$ . Further, suppose that the mutant response rule yields a stable agreement donation of  $u + \delta$  when interacting with a resident individual, where the mutant deviation,  $\delta$ , is small. In this case, at final agreement, the resident individual will make a donation  $u + \Lambda\delta$ , where  $\Lambda$  denotes the overall resident responsiveness (which is yet to be determined). With these specifications, Eq. (1) of main text again holds at the ESS, i.e.,

$$\left[ (1-q)(\Lambda b - c) + q(\hat{\Lambda}b - c) \frac{1-\Lambda}{1-\hat{\Lambda}} \right] \delta + O(\delta^2) < 0. \quad (\text{A.5})$$

We must now calculate the resident responsiveness,  $\Lambda$ .

When a resident interacts with a mutant, the stable agreement values must satisfy  $u + \Lambda\delta = f(u + \delta, u + \Lambda\delta)$ . Expanding this in a Taylor series for small  $\delta$  gives

$$u + \Lambda\delta = f(u, u) + \gamma\delta + \beta\Lambda\delta, \quad (\text{A.6})$$

where  $\beta = \partial f / \partial u_s$  is the resident's responsiveness to self, and  $\gamma = \partial f / \partial u_p$  is its responsiveness to partner. Using the fact that  $f(u, u) = u$ , Eq. (A6) can be solved to give the overall responsiveness,  $\Lambda = \gamma / (1 - \beta)$ .

With these specifications, analyses analogous to those of the main text show that the prevailing response rules satisfy the following conditions:

$$\begin{aligned} \frac{\gamma}{1 - \beta} &= 1, \\ b - c &= 0, \end{aligned} \quad (\text{A.7})$$

which are analogous to the conditions obtained in the simple case of univariate response rules ( $\lambda = 1$  and  $b - c = 0$ , see main text). In the bivariate case as well, prevailing response rules end up mirroring their opponent ( $\gamma / (1 - \beta) = 1$ ), and maximizing the mutual payoff of encounters.

The simplest rule satisfying conditions (A7) is linear, and we refer to it as a generalized reciprocator. From the first condition in Eq. (A7) we have  $\gamma = 1 - \beta$ . Further, the intercept must be zero for any stable agreement to be reached, and therefore the rule is defined as:

$$u'_s = u_s + \gamma(u_p - u_s), \quad (\text{A.8})$$

where the partner-responsiveness  $\gamma$  can take any strictly positive value between zero and one. Thanks to the fact that generalized reciprocators rely on the divergence with partner in order to adjust their own donation, they are guaranteed to play ultimately like any alternative strategy they encounter. As a result, generalized reciprocators not only resist invasion by weak-effect mutants, but resist invasion by any mutant. Furthermore, in a population fixed for any other strategy, generalized reciprocators rise until fixation when starting in any frequency.

#### A.3.1. Payoff-responding rules

The above framework can also be used to analyze the case of payoff-responding rules (Killingback and Doebeli, 2002). These are special cases of the general bivariate response rule  $u'_s = f(u_s, u_p)$  where  $f(u_s, u_p) \equiv g(B(u_p) - C(u_s))$  for some function  $g(I)$  (and corresponding mutant function  $\hat{g}(I)$ ), i.e., individuals respond exclusively to their past payoff.

In this case, the above total responsiveness,  $A$  is restricted to a form where  $\beta = -dg/dI dC/du$ , and  $\gamma = dg/dI dB/du$ . As a result, conditions (A7) require that

$$\frac{dg}{dI} \cdot (b - c) = 1, \quad (A.9)$$

$$b - c = 0.$$

These two conditions cannot hold simultaneously. Consequently, any response rule that responds solely to the payoff in the previous round will always be susceptible to invasion by some weak-effect mutants.

#### A.4. Previous studies of the iterated prisoner's dilemma

##### A.4.1. Analogies with the discrete prisoner's dilemma

In the main text, we show that most evolutionarily stable strategies can be invaded by mutants that reach a large payoff against each other. This mechanism has an equivalent in the simpler case of the discrete iterated prisoner's dilemma (IPD). Unconditional defection is an ESS in the discrete IPD as no other strategy is able to do better than defectors in a population of defectors. However, a strategy called tit-for-tat whereby individuals start by cooperating and then express, in each round, the action expressed by their partner in the previous round, is nevertheless able to invade when it is initially present at high enough frequency, because tit-for-tat individuals reach a very large payoff when playing against each other (Axelrod, 1984; Nowak and Sigmund, 1992; Nowak et al., 2004).

##### A.4.2. Simulation results in the continuous prisoner's dilemma

In a previous work (Wahl and Nowak, 1999a,b), stochastic simulations did not seem to yield the general emergence and fixation of perfect reciprocity, although cooperative strategies with responsiveness close to one had

a remarkable importance (see Fig. 7 of Wahl and Nowak, 1999a). This appears to contradict our analytical results. Although direct comparison of published simulation results with our analysis is difficult, we believe that the resolution stems from subtle aspects of how simulations of continuous IPD are necessarily conducted. In short, we suggest that perfect reciprocity will ultimately prevail in stochastic simulations once it appears (as our analytical results predict and our simulations show; Fig. 2 of main text), but that its appearance is extremely unlikely in many simulation techniques. In the following we describe the general difficulty of observing the evolution of global reciprocity in simulations, and explain why previous works did not observe it.

In the continuous IPD, because there is a continuum of possible strategies, simulations must typically employ mutation to allow for the potential introduction of all of these. The mutation scheme employed in simulations can have significant effects on the outcome of simulations for the following reasons: (i) selection in favor of global reciprocity is weak because it involves rare mutant–mutant interactions. Furthermore, directional selection towards larger  $\lambda$  is typically very weak. From a detailed analysis of the invasion conditions (Eq. (1) of main text), one can show that the mutants that are the most likely to invade are those with  $\lambda$  close to 1 (and with the appropriate intercept), but other mutants with lower  $\lambda$  can nevertheless always invade as well. Therefore, whenever the resident's  $\lambda$  is not strictly equal to one, evolution can occur in both directions, making the outcome highly dependent on the mutation scheme used in the simulations. This can cause the population to evolve somewhat erratically, with little directional tendency. (ii) Related to point (i), if mutation is completely random within strategy space, then the likelihood of global reciprocity appearing is virtually zero. Thus, even though perfect reciprocity would spread to fixation once it appears, this rarely happens and thus the population evolves somewhat erratically. (iii) Related to point (ii), in simulations, the response rules used are always linear. Therefore, when the slope of a response rule is close to  $\lambda = 1$ , the stable agreement reached by individuals playing this rule is extremely sensitive to the intercept of that rule. As a result, virtually none of the mutants that appear with  $\lambda$  close to one will be advantageous.

Given these explanations, the simulation method employed in previous studies (Wahl and Nowak, 1999a,b) could not lead to the definitive fixation of global reciprocity but only to a slight tendency towards large values of  $\lambda$ . Our simulations are different on two respects. First, the criterion that we use for mutant invasion is more restricted and generates more stability. Second, and most importantly, global reciprocity ultimately prevails in our simulations because it is introduced 'by hand' with a given probability in each turn (see Appendix A.2, and Fig. 2).

Given the above considerations, it might be argued that, although perfect reciprocity is the only response rule that can resist invasion by any mutant, this finding is largely

theoretical and of little relevance to real populations. Instead, real population might be expected to evolve in ways more akin to previous simulations (Wahl and Nowak, 1999a,b) than to the analytical results presented here. On the contrary, however, we suggest that the assumptions needed to perform simulations constrain these results in unintended ways. Mainly, simulations assume that response rules are linear and that their properties evolve on a continuous multi-dimensional space. In consequence, global reciprocity is very unlikely to appear. In contrast, we suggest that global reciprocity is likely to arise by mutation far more often than most other strategies because it is very simple on biological terms. And once it does, our results clearly reveal that it will spread to fixation. Even in the discrete IPD where the ‘probability of appearance’ problem is less of an issue because the strategy space is much smaller, when one considers a large range of possible strategies, tit-for-tat must also be introduced ‘by-hand’ in order to evolve (Nowak and Sigmund, 1992). Therefore, although clearly neither simulations nor the present analytical results are complete descriptions of the evolution of any natural biological population, there are good reasons to believe that the insights provided by the analytical results might more closely match such populations.

#### A.4.3. Payoff-responding strategies in the continuous prisoner’s dilemma

Previous results have also revealed that payoff-responding rules can result in the evolution of cooperative behavior (Killingback and Doebeli, 2002), despite the fact that our analysis suggests that there are no strategies in such games that are resistant to invasion by all mutants. This contradiction is only apparent, however, since our results for payoff-responding strategies reveal that there is no single strategy that is able to resist invasion by all weak-effect mutants. This does not imply, however, that some sort of complex polymorphism cannot stabilize as a result of the combined forces of mutation and selection. Indeed, our analysis agrees exactly with the treatment of Killingback and Doebeli (2002) for payoff-responding strategies.

These authors considered the sub-case, analogous to our treatment of payoff-responding strategies, in which the function  $g$  is linear:  $g(I) = \alpha + \beta I$  in their notation (Killingback and Doebeli, 2002). In this case, inequality (A5) can be reversed to determine when any non-zero level of donation is expected to evolve. For rare mutants, letting  $q$  go to zero, we require  $Ab - c > 0$ . Evaluating this condition at  $u = 0$ , using the definition  $A = \beta B'(u)/(1 + \beta C'(u))$  (see Appendix A.3), gives the condition:

$$\beta > \frac{C'(0)}{B'(0)^2 - C'(0)^2}.$$

This special case is exactly the threshold theorem of Killingback and Doebeli (2002). When it is satisfied, some level of cooperative behavior is expected to evolve. In such cases, however, unlike response rules that are functions

only of one’s partner’s previous action, there is no single strategy that can resist any mutant invasion. Rather, the expected outcome of evolution cannot be predicted analytically. Either non-equilibrium behavior results, or an equilibrium involving a polymorphism occurs. Simulation results from Killingback and Doebeli (2002) suggest the latter, but it is not possible to unambiguously distinguish this possibility from very slowly changing non-equilibrium dynamics using their results.

#### A.5. The consequences of errors

Our results on the necessary properties of prevailing response rules are valid for linear as well as non-linear response rules. Whether they are linear or not, the only response rules that can resist invasion by any mutant have the same outcome in terms of (i) cooperation level and (ii) local responsiveness to self and partner. These conditions concerning local properties of response rules around agreement are necessary for resistance to weak-effect mutants (i.e., mutants leading to a stable agreement close to resident’s, see Fig. 1c of main text). However, in order for response rules to resist invasion by any mutant, these conditions must be satisfied globally as well, i.e., response rules must be linear. For instance, in the case of partner-response only, local stability stipulates that response rules must cross the diagonal with a slope equal to one (i.e., they must be tangent to the diagonal), whereas global stability stipulates that the response rule must be linear with a slope equal to one (i.e., the rule must be the diagonal itself, Fig. A2a). Both perfect reciprocation and general reciprocation are globally stable response rules, and they are linear.

However, these linear rules raise another problem. Neither perfect reciprocation nor general reciprocation encompasses any force bringing back donation to the optimal level  $u^*$  would it diverge from this level by mistake. If noise is an important issue, and if interactions really last an infinite amount of time, then donation will certainly drift far away from the optimum. The only way for response rules to maintain their donation forever close to the optimum is to be somehow non-linear. For instance, in the case of partner-response only (but the same reasoning can be employed for bivariate response rules), response rules must over-respond to partner when donation is too low in order to move donation up, and vice versa when donation is too high (Fig. A2b–d). Therefore, as a flip side, these response rules can be exploited by strong effect defectors, benefiting from their over-response at low partner donation. Unfortunately, our model does not allow predicting in a quantitative way the evolution of response rules in this case.

##### A.5.1. Slowing down the drift with inertia

The drift of donation due to the accumulation of errors can be significantly slowed down in the linear case (but not prevented). This occurs with generalized reciprocation and

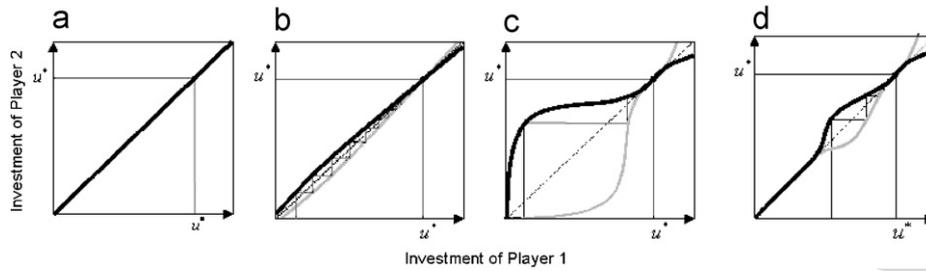


Fig. A2. Examples of response rules resisting invasion by any weak-effect mutant. Every response rule is tangent to the diagonal at the point where donation is optimum ( $u^*$ ). (a) Perfect reciprocation is linear and is therefore the diagonal itself. (b)–(d), Non-linear rules over-respond when partner donates less than  $u^*$  and lower-respond in the opposite case, at least in the neighborhood of  $u^*$ . In the opposite situation, the agreement is unstable. In (d), convergence is possible only if both players make their first offer above a certain threshold.

low partner responsiveness (see Eq. (A8)), as we shall explain in the following. When two identical generalized reciprocators interact, starting, respectively, with donations  $u_1$  and  $u_2$ , then the average donation in any round  $m$  is constant and given by

$$\bar{u}(m) = \frac{u_1 + u_2}{2}.$$

As any linear rule, generalized reciprocation does not contain any directional force affecting average donation.

Let us now consider the following scenario. Two players with the same generalized reciprocating response rule interact. In each round, they both aim at donating  $u^*$ , and they do so until an error occurs. In a round, arbitrarily called 0, player 2 is willing to donate  $u^*$  but donates, or seems to donate,  $u^* + \varepsilon$  instead. In fact, what player 2 really donates does not matter. What matters is the subjective measure of her donation by her partner, as well as by herself. In order to integrate various possibilities in a general framework, let us consider that player 1 measures player 2's donation as  $u^* + \varepsilon$ , while player 2 thinks that she has donated  $u^* + s \cdot \varepsilon$ , where  $s \in [0, 1]$ . A low  $s$  means that player 2 is better than player 1 at measuring accurately her own donation or at knowing her own intention. In most cases,  $s$  should be 0 because player 2 firmly knows that she donated or was willing to donate  $u^*$ .

In the next round, player 1 responds to what she observes by playing  $u_1 = u^* + \gamma \cdot \varepsilon$ , whereas player 2 responds by playing  $u_2 = u^* + s \cdot \varepsilon(1 - \gamma)$ . Therefore, in the following of their interaction and in the absence of ulterior errors, the average donation of both players in any round will be

$$\bar{u} = u^* + \frac{\varepsilon}{2} [s + (1 - s)\gamma].$$

As long as individuals have a superior access to their own behavior or intentions than other individuals (i.e.,  $s < 1$ ), then the average cost of error in each round increases with the level of responsiveness to partner ( $\gamma$ ) (see discussion in main text).

This result can be readily extended to consider the case where errors are being made in each round. Because generalized reciprocation does not encompass any direc-

tional force, the average donation in each round is constant except for errors. Therefore after  $n$  rounds with an error  $\varepsilon_i$  made in each round, the average donation is

$$\bar{u}(n) = u^* + \frac{s + (1 - s)\gamma}{2} \sum_{i=1}^n \varepsilon_i.$$

Assuming that errors have standard deviation  $\sigma$  and are unbiased (i.e., the expectation of  $\varepsilon_i$  is zero) then the standard deviation of average donation in round  $n$  is given by

$$\bar{\sigma}(n) = \sigma \sqrt{n} \sqrt{\frac{s + (1 - s)\gamma}{2}}.$$

The standard deviation of donation is linearly increasing with the number of rounds because errors accumulate. However, if  $s < 1$ , then the larger is inertia (low  $\gamma$ ), the slower errors accumulate.

In practice, in order to avoid the indefinite accumulation of errors, strategies must necessarily encompass a certain degree of non-linearity. However, it is not possible to follow analytically the successive donations made by each individual when rules are non-linear. Yet, locally around stable agreement, non-linear rules behave like linear ones (see Eq. (A7)). In other words, if non-linear rules are able to bring back donation closer to  $u^*$  when it diverges from it, it is only because of their curvature. Therefore, even though it cannot be demonstrated analytically, it seems plausible that, even with non-linear rules, inertia should slow down the accumulation of errors and thus allow individuals to play generally closer to the optimum. Of course, if social interactions do not really last indefinitely, then such inertia comes at a cost, because it delays one's reaction to potential defectors. Therefore, quantitatively, inertia should be at a balance between the cost of errors and the potential cost of being cheated.

## References

- Axelrod, R., 1984. The Evolution of Cooperation. Basic Books, New York.
- Boyd, R., Lorberbaum, J.P., 1987. No pure strategy is evolutionarily stable in the repeated prisoners-dilemma game. Nature 327, 58–59.

- Doebeli, M., Hauert, C., 2005. Models of cooperation based on the prisoner's dilemma and the snowdrift game. *Ecol. Lett.* 8, 748–766.
- Hamilton, W.D., 1964. The genetical evolution of social behavior, I. *J. Theor. Biol.* 7, 1–52.
- Killingback, T., Doebeli, M., 2002. The continuous prisoner's dilemma and the evolution of cooperation through reciprocal altruism with variable investment. *Am. Nat.* 160, 421–438.
- Maynard Smith, J., Price, G.R., 1973. Logic of animal conflict. *Nature* 246, 15–18.
- McNamara, J., Gasson, C., Houston, A., 1999. Incorporating rules for responding into evolutionary games. *Nature* 401, 368–371.
- Nowak, M., Sigmund, K., 1992. Tit for tat in heterogeneous populations. *Nature* 355, 250–253.
- Nowak, M., Sigmund, K., 1993. A strategy, of win stay lose shift that outperforms tit-for-tat in the prisoners-dilemma game. *Nature* 364, 56–58.
- Nowak, M.A., Sasaki, A., Taylor, C., Fudenberg, D., 2004. Emergence of cooperation and evolutionary stability in finite populations. *Nature* 428, 646–650.
- Proulx, S., Day, T., 2001. What can invasion analyses tell us about evolution under stochasticity in finite populations? *Selection* 2, 1–15.
- Rousset, F., 2003. A minimal derivation of convergence stability measures. *J. Theor. Biol.* 221, 665–668.
- Rousset, F., Billiard, S., 2000. A theoretical basis for measures of kin selection in subdivided populations: finite populations and localized dispersal. *J. Evol. Biol.* 13, 814–825.
- Taylor, P.D., Day, T., 2004. Stability in negotiation games and the emergence of cooperation. *Proc. R. Soc. London B* 271, 669–674.
- Taylor, P.D., Frank, S.A., 1996. How to make a kin selection model. *J. Theor. Biol.* 180, 27–37.
- Trivers, R., 1971. The evolution of reciprocal altruism. *Q. Rev. Biol.* 46, 35–57.
- Wild, G., Taylor, P., 2004. Fitness and evolutionary stability in game theoretic models of finite populations. *Proc. R. Soc. London B* 271, 2345–2349.
- Wahl, L.M., Nowak, M.A., 1999a. The continuous prisoner's dilemma: I. Linear reactive strategies. *J. Theor. Biol.* 200, 307–321.
- Wahl, L.M., Nowak, M.A., 1999b. The continuous prisoner's dilemma II. Linear reactive strategies with noise. *J. Theor. Biol.* 200, 323–338.