vol. 185, no. 3 the american naturalist march 2015

# Contingency in the Evolutionary Emergence of Reciprocal Cooperation

**Jean-Baptiste André***

Centre National de la Recherche Scientifique, Institut de Biologie de l'Ecole Normale Supérieure, Unité Mixte de Recherche 8197, 46 rue d'Ulm, 75005 Paris, France

*Submitted February 10, 2014; Accepted November 6, 2014; Electronically published January 28, 2015*

*Online enhancements:* appendix, zip file.

abstract: Reciprocity is characterized by individuals actively making it beneficial for others to cooperate by responding to them. This makes it a particularly powerful generator of mutual interest, because the benefits accrued by an individual can be redistributed to another. However, reciprocity is a composite biological function, entailing at least two subfunctions: (i) a behavioral ability to provide fitness benefits to others and (ii) a cognitive ability to evaluate the benefits received from others. For reciprocity to evolve, these two subfunctions must appear together, which raises an evolutionary problem of bootstrapping. In this article, I develop mathematical models to study the necessary conditions for the gradual emergence of reciprocity in spite of this bootstrapping problem. I show that the evolution of reciprocity is based on three conditions. First, there must be some variability in behavior. Second, cooperation must pre-evolve for reasons independent of reciprocity. Third, and most significantly, selection favors conditional cooperation only if the cooperation expressed by others is already conditional, that is, if some reciprocity is already present in the first place. In the discussion, I show that these three conditions help explain the specific features of the instances in which reciprocity does occur in the wild. For instance, it accounts for the role of spatial symmetry (as in ungulate allogrooming), the importance of synergistic benefits (as in nuptial gifts), the facilitating role of collective actions (as in many instances of human cooperation), and the potential role of kinship (as in primate grooming).

*Keywords:* evolution of cooperation, reciprocity, mechanistic constraints, bootstrapping.

## Introduction

Many cooperative traits, particularly, though not only, in humans, are expressed toward nongenetically related partners, in which case, they must be mutualistic (West et al. 2007). A potentially important mechanism by which two individuals can mutually benefit from helping each other is reciprocity (Trivers 1971), which in a broad sense characterizes a variety of mechanisms of social feedback, including the reciprocal exchange of episodes of help between two partners (called direct reciprocity), the effect of reputation on partner choice, and punishment (also called negative reciprocity). However, reciprocity is also the subject of an evolutionary puzzle: a discrepancy between theoretical predictions and empirical observations. On one hand, evolutionary models show that reciprocity can evolve relatively easily, provided there are repeated encounters between players (Axelrod and Hamilton 1981; see also, e.g., Nowak and Sigmund 1992, 1993; Roberts and Sherratt 1998; Lehmann and Keller 2006; André and Day 2007). On the other hand, relatively few instances of reciprocity have been empirically demonstrated in nonhuman animals, and the interpretation of empirical observations is subject to intense debate (Connor 1986, 1995*a*, 1995*b*; Hammerstein 2003; Bergmuller et al. 2007; West et al. 2007; Clutton-Brock 2009; Leimar and Hammerstein 2010).

In a recent article, I suggested that one possible reason for the relative rarity of reciprocity is that it raises an evolutionary problem of bootstrapping, which makes it very unlikely to evolve away from defection (André 2014). It has long been known that the evolution of reciprocity poses a specific problem due to the fact that it is favored only if a sufficient proportion of other individuals in the population already reciprocate. Hence, the evolutionary emergence of reciprocity requires the crossing of an invasion barrier. Theorists, however, have given the impression that this barrier could be crossed relatively easily through genetic drift and/or large mutation rates (Nowak and Sigmund 1992, 1993, 1994, 1995; Kandori et al. 1993; Hauert and Schuster 1997; Brauchli et al. 1999; Hauert and Stenull 2002; McNamara et al. 2004; Nowak et al. 2004; Imhof et al. 2005; Imhof and Nowak 2010). Yet this solution is premised on a very strong, albeit generally unacknowledged mechanistic assumption: that reciprocal co-

---

* Present address: Institute of Evolutionary Sciences, CC 065, Place Eugène Bataillon, Montpellier 34095, France; e-mail: jeanbaptisteandre@gmail.com.

Am. Nat. 2015. Vol. 185, pp. 303–316.
DOI: 10.1086/679625

operation is a simple biological function, able to emerge out of defection through a single (or very few) random mutations that can then drift neutrally, eventually crossing the invasion barrier. In reality, however, there is no reason why this should be the case. To the contrary, the ability to cooperate reciprocally, like any biological function (see Orr 2005 for a review), is likely to be a composite trait entailing several adaptive mutations (see Stevens and Hauser 2004; Stevens et al. 2005). In particular, reciprocity involves at least two different functions: (i) the behavioral ability to provide benefits to others (i.e., to help) and (ii) the cognitive ability to detect (and respond to) the benefits provided by others. In evolution, composite functions are normally shaped by the enduring effect of natural selection, which allows the accumulation of adaptive mutations. In the case of reciprocity, however, except under specific biological conditions (see below), this gradual accumulation is impossible since the selective pressure favoring reciprocity is present only once reciprocity is in place. As a result, in contrast to other composite functions, the ability to reciprocate cannot build up gradually. It would need to arise by chance (i.e., by the occurrence of just the right mutations) and then become favored by selection. André (2014) showed that this is highly unlikely.

This leads to the opposite puzzle, however, which is the subject of this article. Reciprocal cooperation probably does exist in a number of cases (see Raihani and Bshary 2011 for a review), and there is no doubt of its existence in humans. Moreover, if we consider cooperation evolved by partner choice as a form of reciprocity in the broad sense, it can even be argued that reciprocity is not really so rare, as partner choice is likely taking place in many interactions in nature (Noë et al. 2001). Hence, reciprocal cooperation must sometimes be able to evolve, in spite of the fact that it generally raises a bootstrapping problem. The aim of this article is to characterize, with the help of mathematical models, the biological conditions that facilitate this evolution. Beyond explaining the rarity of reciprocity per se (which can always be debated), this analysis will prove useful in explaining the precise form that reciprocity takes in extant species.

Theorists have already attempted to model the gradual evolution of reciprocal cooperation under the assumption that it requires the accumulation of several mutations (Lehmann and Keller 2006; André and Day 2007; Akçay et al. 2009). But in doing so they have made assumptions that facilitate the task and undermine their generality. In Akçay et al. (2009), a single quantitative trait called "other-regarding preferences" is assumed to cause both (i) cooperation in general and (ii) the ability to respond to a partner's cooperation. Lehmann and Keller (2006) consider two heritable traits, but these two traits are (i) the tendency to cooperate in the first round of an interaction and (ii) responsiveness to a partner's cooperation in all subsequent rounds; hence, in all rounds except the first, cooperating and responding to a partner are caused by the same genetic trait. In both cases (Lehmann and Keller 2006; Akçay et al. 2009), therefore, specific mechanistic assumptions regard selection for reciprocity as an automatic by-product of selection for constitutive cooperation. In André and Day (2007), we did not assume particularly facilitating mechanisms of this kind, but we found that selection for responsive cooperation was only a second-order force, which played a significant role only because we assumed the absence of first-order effects (e.g., the slightest cognitive cost of conditionality would have prevented the evolution of reciprocity in the model presented here). All of these assumptions indeed facilitate the evolution of reciprocity but have no reason to be general.

This study aims to account for the gradual emergence of reciprocal cooperation under more general assumptions. Because the evolution of reciprocity poses a bootstrapping problem, it depends crucially on assumptions regarding the biological mechanisms underlying social behavior (note that Akçay et al. 2009 already observed the importance of mechanisms in the evolution of reciprocity). The problem, however, is that all possible mechanistic assumptions cannot be considered in a single model. It is therefore tempting to conclude that the evolution of reciprocity is a case-by-case issue with no general principles. One of the findings of this work, however, is that this is not true and that one can identify general properties of biological situations that may facilitate the emergence of reciprocity.

I first develop a general model aimed at measuring the selective pressure acting respectively on the two ingredients of reciprocal cooperation: the ability to cooperate and the ability to condition one's cooperation on the cooperation of others. I then consider two forms of social interaction, (1) under the assumption that cooperation is all or nothing (but probabilistic) and (2) under the assumption that cooperation can take on a range of values, but always assuming weak selection. I show that the same result holds in all cases: namely, that the evolution of reciprocity requires that some form of conditional cooperation preexists for an independent reason. I then discuss the major mechanisms that can play such a triggering role and show, through examples, that they do account for the forms that reciprocal cooperation takes in extant species.

## A General Model

To disentangle the various selective pressures acting on reciprocal cooperation, I first develop a general argument (an even more general version of the same argument can be found in the appendix, available online). Consider an interaction between two individuals, lasting for any length of time. The interaction is made up of a succession of rounds, each consisting of the simultaneous expression of various amounts of cooperation by both partners. Note that this cooperation may take the form of either expressing a helping action or refraining from a harming action. In both cases, this cooperation involves a personal cost to the actor (at least temporarily) and a benefit to the recipient.

Consider a trait $z$ affecting the social strategy played by individuals in an unspecified way (see table 1 for a list of the parameters for this model). I consider a focal player with trait $z_\bullet$ with a partner with trait $z_0$. For the sake of simplicity, the effect of genetic relatedness is not considered in the main text (but see "Inclusive Fitness in the General Approach" in the appendix). Moreover, because I am looking at the first-order effect of small variations of $z$, I measure only the direction of selection on $z$ stemming from its unitary effect in a single round, arbitrarily called round 0.

In round 0, the two individuals have information about the history of their interaction, that is, the series of cooperative decisions made by each individual in the past. The direction of selection on $z$ may depend on its effect after various such histories (e.g., $z$ may stipulate more cooperation after the partner has been generous but less cooperation otherwise). Hence, to measure the selection on $z$, we must average its effect on all possible histories. Let us label any given history of the interaction before round 0 as $\eta$, chosen from a random vector (a multivariate random variable) of unspecified distribution $H$, which represents the distribution of all possible histories of an interaction before round 0.

The amount of cooperation expressed by the focal individual in round 0 after $\eta$ is called $h_\bullet^0$, and the total amounts of cooperation expressed by the focal individual and its partner from round 0 (included) until the end of the interaction are called $h_\bullet^T$ and $h_0^T$, respectively. The focal individual's fecundity is then assumed to be a linear function of cooperation, given by $F_\bullet = bh_0^T - ch_\bullet^T$, where $b$ and $c$ are the unitary benefit and cost of cooperation, respectively. The marginal effect of $z$ on fecundity after history $\eta$ is then written as $\partial F_\bullet/\partial z_\bullet = b\partial h_0^T/\partial z_\bullet - c\partial h_\bullet^T/\partial z_\bullet$. From the chain rule, this yields

$$\frac{\partial F_\bullet}{\partial z_\bullet} = b\frac{\partial h_0^T}{\partial h_\bullet^0}\frac{\partial h_\bullet^0}{\partial z_\bullet} - c\frac{\partial h_\bullet^T}{\partial h_\bullet^0}\frac{\partial h_\bullet^0}{\partial z_\bullet}. \tag{1}$$

We now need to consider the fact that the eventual effect of the trait $z$ on the focal's fecundity depends on the effect of $z$ in many different contexts, that is, after many possible histories. Hence, we need to average equation (1) over the distribution $H$ of all possible histories before 0. For any property $x$ of the interaction, call $\mathrm{E}[x]$ the expectation of $x$ over $H$.

**Table 1:** Main parameters of the general model

| Parameter | Definition |
|---|---|
| $z_i$ | Trait value of individual $i$ ($i = \bullet$ is the focal individual, and $i = 0$ is his partner) |
| $\eta$ | Given history of the interaction before round 0 |
| $H$ | Distribution of all possible histories before 0 |
| $h_i^0$ | Helping level of individual $i$ in round 0 |
| $h_i^T$ | Total helping expressed by individual $i$ from round 0 until the end of the interaction |
| $F_i$ | Fecundity of individual $i$ |
| $b$ | Linear benefit of receiving help |
| $c$ | Linear cost of providing help |
| $\lambda$ | Marginal effect of $z_i$ on individual $i$'s own amount of help in round 0 |
| $\rho$ | Responsiveness of $i$'s partner to the amount of help expressed by $i$ in round 0 |
| $\sigma$ | Responsiveness of $i$ to his own amount of help in round 0 |

Let me now define $\lambda \equiv \partial h_\bullet^0/\partial z_\bullet$ as the marginal effect of $z$ on one's own amount of cooperation expressed in round 0; hence, $\mathrm{E}[\lambda]$ represents the average effect of $z$ on cooperation in round 0. Let me also define $\rho \equiv \partial h_0^T/\partial h_\bullet^0$ as the partner's total responsiveness to the amount of cooperation expressed by the focal in round 0 and $\sigma \equiv \partial h_\bullet^T/\partial h_\bullet^0$ as the focal individual's total responsiveness to his own amount of cooperation in round 0.

From equation (1), averaged over the distribution $H$ of all possible histories, the direction of selection on $z$ is then proportional to a selection gradient $S \equiv \mathrm{E}[\lambda(b\rho - c\sigma)]$, which becomes:

$$\begin{aligned} S = {} & \mathrm{E}[\lambda] \cdot \mathrm{E}[b\rho - c\sigma] + b \cdot \mathrm{Cov}(\lambda, \rho) \\ & - c \cdot \mathrm{Cov}(\lambda, \sigma). \end{aligned} \tag{2}$$

Equation (2) has three terms with interesting interpretations. The first, $\mathrm{E}[\lambda] \cdot \mathrm{E}[b\rho - c\sigma]$, measures the direction of selection on the expected amount of cooperation (i.e., the effect of having $\mathrm{E}[\lambda] > 0$), which is positive if the overall effect of increased cooperation is positive. The second and third terms of equation (2), $b \cdot \mathrm{Cov}(\lambda, \rho) - c \cdot \mathrm{Cov}(\lambda, \sigma)$, measure the benefit of improving the conditionality of cooperation per se independently of its expected level. The second term measures the effect of conditionality on the social benefit of cooperation, whereas the third measures the effect of conditionality on its individual cost.

To understand intuitively, assume that $z$ affects the circumstances under which one cooperates in round 0 (e.g., one cooperates more with partners who have been highly cooperative in the past but less with others) but not one's average cooperativeness (hence, $\mathrm{E}[\lambda] = 0$), and assume further that the individual cost of cooperation is independent of circumstances ($\mathrm{Cov}(\lambda, \sigma) = 0$). In this case, $z$ can be favored if $\mathrm{Cov}(\lambda, \rho) > 0$, that is, if it stipulates to cooperate more ($\lambda > 0$) at histories after which it turns out that the partner will respond more positively to cooperation (i.e., in circumstances in which $\rho$ happens to be large). In other words, $z$ can be favored if it stipulates to cooperate more (less), not only with more (less) generous partners but, more precisely, with more (less) responsive partners because cooperation is more (less) worthwhile with them. In the appendix, I show that the same principle also holds in a more complex framework in which individuals can choose to allocate their cooperative effort to various partners (i.e., partner choice is possible).

The bottom line of this general argument is that selection for responsiveness per se is positive only if partners already respond more or less generously to cooperation because then each unit of investment spent with them is more or less profitable and it makes sense to adapt one's own level of cooperation to theirs. Hence, the evolution of responsiveness depends on the preexistence of responsiveness. In what follows, we will see that this has important consequences for our understanding of the origin of reciprocal cooperation.

## Mechanistic Models

In order to better understand the joint evolution of cooperation and conditionality, I now look at more specific models in which the mechanistic effects of individuals' strategies are considered explicitly. The notation that follows is entirely independent of the general argument above (see table 2 for a list of parameters for this second model).

Consider a pairwise quantitative social interaction in which each partner $i$ invests a total amount $h_i$ in helping the other. I later consider the fact that $h_i$ may be causally determined by microscopic behavioral traits expressed by each partner in various ways, but for now I treat this point at a general level. After an interaction in which a focal individual cooperates a total amount $h_{\bullet}$ and the partner cooperates a total amount $h_0$, I assume that the social payoff gained by the focal individual is

$$P(h_{\bullet}, h_0) = ah_{\bullet} + bh_0 - c(h_{\bullet})^2. \tag{3}$$

In this equation, $ah_{\bullet}$ represents the automatic individual benefit of helping that accrues to the helper owing to common interest with the helpee, $bh_0$ represents the social benefit of receiving help, and $c(h_{\bullet})^2$ is the cost of helping, which increases more than linearly with investment, thereby leading to an optimal intermediate level of investment.

**Table 2:** Main parameters of the mechanistic model

| Parameter | Definition |
|---|---|
| $h_i$ | Total helping expressed by individual $i$ |
| $a$ | Linear benefit of providing help (due to a common interest between partners) |
| $b$ | Linear benefit of receiving help |
| $c$ | Coefficient of the quadratic cost of providing help |
| $\boldsymbol{\tau}_i$ | Vector of microscopic traits of individual $i$ |
| $\sigma_z^2$ | Phenotypic variance in the expression of trait $z$ |
| $F(\boldsymbol{\tau}_i, \boldsymbol{\tau}_j)$ | Fecundity of an individual expressing $\boldsymbol{\tau}_i$ when his partner expresses $\boldsymbol{\tau}_j$ |
| $\tilde{F}(\boldsymbol{\tau}_i, \boldsymbol{\tau}_j)$ | Expected fecundity of an individual with genetic value $\boldsymbol{\tau}_i$ with a partner with $\boldsymbol{\tau}_j$ |
| $R$ | Average genetic relatedness between partners |
| $S_z$ | Measure of the gradient of selection on a trait $z$ |
| $\gamma_i$ | Cooperativeness of individual $i$ |
| $\rho_i$ | Degree of conditionality of individual $i$ |
| $k$ | Linear cost of conditional abilities |

### *Microscopic Traits*

Based on the framework described above, I now consider two different models of interaction, in which the amount of helping expressed by each partner is determined by microscopic behavioral traits. I present only the first of these models in the main text, the other model is presented in the appendix ("Alternative Microscopic Models"). For now, in order to remain general, I consider that each individual is characterized by a vector $\boldsymbol{\tau} = \{\tau^i, i \in [\![1, n]\!]\}$ of $n$ microscopic traits. Hence, the total amount of helping offered by a focal player in an interaction with a partner is $h(\boldsymbol{\tau}_{\bullet}, \boldsymbol{\tau}_0)$, where $\boldsymbol{\tau}_{\bullet}$ and $\boldsymbol{\tau}_0$ represent the vectors of microscopic traits of the focal individual and the partner, respectively. The focal individual's fecundity is then

$$F(\boldsymbol{\tau}_{\bullet}, \boldsymbol{\tau}_0) = P(h_{\bullet}, h_0) - K(\boldsymbol{\tau}_{\bullet}), \tag{4}$$

where $P(h_{\bullet}, h_0)$ is the social payoff as given by equation (3), with $h_{\bullet} = h(\boldsymbol{\tau}_{\bullet}, \boldsymbol{\tau}_0)$ and $h_0 = h(\boldsymbol{\tau}_0, \boldsymbol{\tau}_{\bullet})$, and $K(\boldsymbol{\tau})$ measures the physiological cost of the strategy $\boldsymbol{\tau}$. This cost is included to account for the fact that conditional strategies are likely to be more costly than constitutive ones.

### *Introducing Variability in a Simple Way*

As the general model of the first section has helped clarify, selective pressures in favor of conditionality stem from the presence of some variability in partners' behavior (see McNamara and Leimar 2010 for a review of this idea). However, introducing such variability in a model can easily make it intractable. A first possibility would be to consider higher-order effects of genetic variance and thus abandon the weak selection assumption, but this would complicate dramatically the analysis and require the extensive use of simulations (e.g., as in McNamara et al. 2008). This would not allow clear disentanglement of the selective forces at work.

Therefore, in order to capture the effect of variability in the simplest possible way, I consider the effect of phenotypic, rather than genotypic, variance in the expression of microscopic traits. I consider the fact that, for unspecified reasons, one's partner may vary in the expression of underlying genetic traits throughout the course of an interaction. The biological idea behind this approach is that each individual is characterized by a nonheritable state variable representing, for instance, the payoff received from cooperating (see Leimar 1997; André 2010) and responds plastically in social behavior to the value of this state variable. Hence, the variability of the state variable yields variability in social behavior. However, to keep things as simple as possible, I do not consider such a state variable explicitly in the main text (but see the appendix). I simply assume that, in any given interaction, an individual with genetic value $\boldsymbol{\tau}$ consistently expresses a slightly modified vector of traits: $\boldsymbol{\tau} + \boldsymbol{\delta}$, where $\boldsymbol{\delta} = \{\delta^i, i \in [\![1, n]\!]\}$ is a vector of values sampled into $n$ independent centered random variables. In this way, the past behavior of the partner contains some information about the partner's actual level of expression of social traits, information that may be worth responding to.

In principle, however, when deriving the fitness of an individual, the fact that the individual's own phenotypic traits may stochastically vary around his genetic value should also be considered. As a result, selection could favor responding to a partner because its behavior indirectly conveys some information on the expression of one's own phenotypic traits. This would be an odd, artifactual consequence of the model. For this reason, in the derivation of a focal individual's fitness, I assume that the individual has the ability to perfectly control the expression of his own social strategy, and I consider only the effect of variability in the partner's behavior. The rationale behind this assumption is, again, in line with the idea that individuals are characterized by an underlying state variable. If behavioral variability is the consequence of the variability of such a state variable, then the focal individual will respond directly to his state variable and not to a partner's behavior as an indirect source of information about it.

As a complementary analysis, however, in the appendix I develop an explicit model in which individuals are characterized by an underlying state variable and can both respond plastically to this variable and/or respond to their partner's behavior. This more complex model yields the same results as the simple model presented here, which allows better insight into the selective pressures at work.

With such phenotypic variance in the partner's behavior, the focal individual's fecundity is now $F(\boldsymbol{\tau}_\bullet, \boldsymbol{\tau}_0 + \boldsymbol{\delta}_0)$. Assuming that the noise is small, this expression can be written to second order in each $\delta_0^i$. Writing $\mathrm{E}[\cdot]$ for the expectation of a random variable over the distribution of phenotypic deviations, the focal individual's expected fecundity $\tilde{F}(\boldsymbol{\tau}_\bullet, \boldsymbol{\tau}_0)$ is given by

$$\begin{aligned}\tilde{F}(\boldsymbol{\tau}_\bullet, \boldsymbol{\tau}_0) &= \mathrm{E}[F(\boldsymbol{\tau}_\bullet, \boldsymbol{\tau}_0 + \boldsymbol{\delta}_0)] \\ &= F(\boldsymbol{\tau}_\bullet, \boldsymbol{\tau}_0) + \sum_i \frac{\sigma_i^2}{2} \frac{\partial^2 F(\boldsymbol{\tau}_\bullet, \boldsymbol{\tau}_0)}{\partial \boldsymbol{\tau}_0^{i2}} + \sum_i o(\sigma_i^2),\end{aligned} \tag{5}$$

where $\partial^2 F(\boldsymbol{\tau}_\bullet, \boldsymbol{\tau}_0)/\partial \boldsymbol{\tau}_0^{i2}$ is the partial derivative of the focal individual's fitness function with respect to the *i*th microscopic trait of the partner, evaluated at the expected value of all microscopic traits, and $\sigma_i^2$ represents the variance due to noise in the expression of the *i*th microscopic trait of the partner. To first order in genetic variance, the linkage disequilibrium between microscopic traits can be neglected, as it yields second-order effects.

The effect of genetic relatedness also needs to be considered. In this regard, and to remain as simple as possible, the model assumes that competition is homogeneous in the global population (i.e., there is no difference in amounts of competition with kin and with nonkin) and that relatedness between social partners is generated by an unspecified assortment process. Hence, the direction of selection on each microscopic trait $\boldsymbol{\tau}^i$ is simply given by the sign of

$$S_i = \frac{\partial \tilde{F}(\boldsymbol{\tau}_\bullet, \boldsymbol{\tau}_0)}{\partial \boldsymbol{\tau}_\bullet^i} + R \frac{\partial \tilde{F}(\boldsymbol{\tau}_\bullet, \boldsymbol{\tau}_0)}{\partial \boldsymbol{\tau}_0^i}, \tag{6}$$

where $R$ is the genetic relatedness between social partners as measured on neutral loci (Rousset 2004).

## Microscopic Models

The above model is quite general and could apply to many forms of social interaction between two partners. I now specify the nature of the microscopic traits and the way they control the behavior of players in social interactions. Here I describe one model adapted from André (2014). I present alternative microscopic models in the appendix, with essentially identical results.

Individuals are haploid and characterized by two heritable microscopic traits: their cooperativeness $\gamma$ and their degree of conditionality $\rho$. Conditional abilities carry a physiological cost $k \times \rho$. Cooperation in each round is 0 or 1, probabilistically (see appendix for a different assumption). Consider an interaction between a focal individual with traits ($\gamma_\bullet$, $\rho_\bullet$) and a partner with traits ($\gamma_0$, $\rho_0$). After the partner has cooperated in the preceding round, the focal individual has a probability $\gamma_\bullet$ of cooperating. After the partner has defected in the preceding round, the focal has a probability $\gamma_\cdot(1-\rho_\cdot)$ of cooperating. Hence, each individual's probability of cooperating in a given round can be calculated, round after round, by recurrence. Assuming that both partners always cooperate in the first round and that the interaction lasts for an infinite length of time, such that the initial nonstationary rounds can be neglected, it can be shown that the total amount of helping offered by the focal individual is proportional to

$$h(\gamma_\cdot, \rho_\cdot, \gamma_0, \rho_0) = \frac{\gamma_\cdot\{1-\rho_\cdot[1-\gamma_0(1-\rho_0)]\}}{1-\gamma_\cdot\gamma_0\rho_\cdot\rho_0}, \quad (7)$$

which is valid provided cooperativeness remains lower than or equal to 1 (i.e., $\gamma \in [0, 1]$) and conditionality remains strictly lower than 1 (i.e., $\rho \in [0, 1[$).

In this model, the variances of both variables ($\gamma$ and $\rho$) turn out to play essentially the same role (this is not the case in the alternative model presented in the appendix). Hence, for simplicity, I assume that only $\gamma$ is subject to stochastic variability with variance $\sigma^2$. The direction of selection on both traits is then found by applying equation (6).

## Results

### Degeneracy of the Condition for Evolutionary Stability

In the absence of phenotypic variability ($\sigma^2 = 0$) and when the cost of conditionality is nil ($k = 0$), simple algebra shows that the direction of selection on $\rho$ can always be expressed as $S_\rho = -\gamma(1-\gamma)S_\gamma/(1-\rho)$, where $S_\gamma$ is the direction of selection on $\gamma$ (a cumbersome expression, not shown). Hence, the joint condition for evolutionary stability on both variables reduces to a single condition. This degeneracy (already observed in different models by Lehmann and Keller 2006; André and Day 2007; Akçay et al. 2009) is a symptom of the fact that, in the absence of behavioral variability in partners, there is no selective pressure acting on conditionality per se (conditionality may, at best, evolve neutrally by genetic drift). Conditionality is selected for only via its effect on the level of cooperation eventually reached. Cooperativeness and conditionality are hence two microscopic traits controlling a single macroscopic outcome, and an infinite number of pairs of traits ($\gamma$, $\rho$) can thus be evolutionarily stable. This can also be observed by plotting the selection gradients acting on $\gamma$ and $\rho$ as a vector field (fig. 1*a*). For instance, when $\rho = 0$, then the corresponding evolutionarily stable cooperativeness is $(a + bR)/2c$, which is typically low if genetic relatedness $R$ and common interest $a$ are both low. In contrast, the maximal level of cooperativeness ($\gamma = 1$) can also be evolutionarily stable if $\rho$ is above a certain threshold (calculable, but not shown).

Things are different when first-order selective pressures acting specifically on conditionality are taken into account. In André and Day (2007), we did not consider such selective pressures, and so only second-order effects mattered. Here I consider the effect of two selective forces acting on conditionality. The first, the cost of conditionality, acts against conditionality. The second, the existence of variability in the partner's behavior, may act in favor of conditionality. The two are introduced one after the other.

### The Cost of Conditionality

With $k > 0$, selection on conditionality can now be written $S_\rho = -k - \gamma(1-\gamma)S_\gamma/(1-\rho)$, where $S_\gamma$ is the direction of selection on $\gamma$. Hence, selection on $\rho$ is always strictly negative when cooperativeness is under positive selection or is evolutionarily stable (i.e., when $S_\gamma \geq 0$). Hence, responsiveness cannot be positive at the condition for evolutionary stability. This can be observed more fully by plotting selection gradients as a vector field (fig. 1*b*). In the absence of variability, there is no point in responding to the partner's behavior, since it is not subject to any uncertainty. Hence, as long as conditionality has a cost, constitutive cooperation is always favored.

### Phenotypic Variability

Let us now consider the effect of introducing some variability in the partner's behavior (i.e., $\sigma^2 > 0$).

*Conditionality Cannot Rise from Zero.* In the initial absence of conditionality ($\rho = 0$), selection on $\rho$ is always $S_\rho = -k - \gamma(1-\gamma)S_\gamma$. Hence, when the population is initially seeded with pure defectors and cooperation rises (i.e., $S_\gamma \geq 0$), selection is always negative on $\rho$; that is, conditionality in cooperation cannot be favored. This can also be observed by plotting the selection gradients as a vector field (fig. 1*d*). Note that $\rho$ may rise by selection if the population is initially seeded with a very high $\gamma$, which then needs to decrease by selection ($S_\gamma < 0$). In this case, conditionality happens to reduce cooperation and is favored only for this reason (these kinds of by-product effects are discussed in more detail in the alternative microscopic models of the appendix).

*But Conditionality Has an Autocatalytic Effect.* Things are different if some conditionality is initially present for an unspecified reason (i.e., if $\rho > 0$; see fig. 1*c*, 1*d*). Mathematically, the direction of selection on $\rho$ when $\rho > 0$ can be expressed as

$$S_\rho = -k - \frac{\gamma(1-\gamma)}{1-\rho}S_\gamma + \sigma^2\gamma\rho \times Q + o(\sigma^2), \quad (8)$$

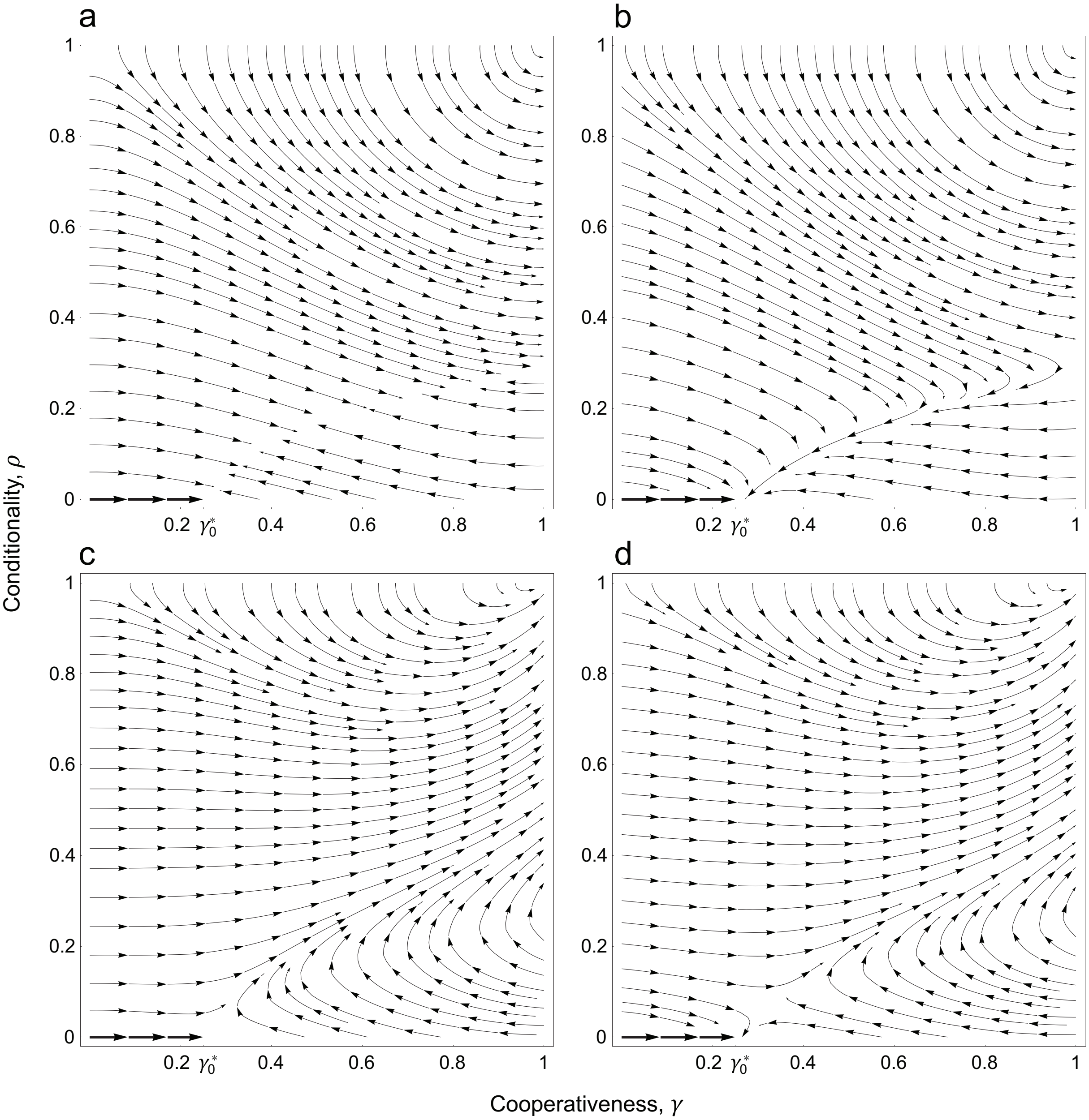


**Figure 1:** Direction of selection on cooperativeness and conditionality. Stream plots showing the direction of selection on $\gamma$ and $\rho$ in the mechanistic model presented in the main text. The vectors in boldface represent the evolutionary trajectory of a population seeded with $\gamma = 0$ and $\rho = 0$. Value $\gamma_0^*$ is the evolutionarily stable cooperativeness in the absence of conditionality ($\rho = 0$). *a*, Conditionality has no cost ($k = 0$), and phenotypic variability is absent ($\sigma^2 = 0$). *b*, Conditionality is costly ($k = 0.1$), but phenotypic variability is absent ($\sigma^2 = 0$). *c*, Phenotypic variability is present ($\sigma^2 = 0.5$), but conditionality has no cost ($k = 0$). *d*, Both phenotypic variability and a cost of conditionality are present ($k = 0.1$ and $\sigma^2 = 0.5$). Other parameters are $a = 1$, $b = 10$, $c = 2$, and $R = 0$.

where $Q$ is a cumbersome expression shown in the appendix ("Direction of Selection in the Microscopic Model of the Main Text"), which is defined (and thus finite) when $\gamma = 0$ and/or $\rho = 0$, and which can be positive and growing with $\gamma$ and $\rho$ (although no formal analysis of this expression was performed). Hence, equation (8) shows that, in the course of the evolution of cooperation (i.e., when $S_\gamma \geq 0$), selection can be positive on conditionality only provided some conditional cooperation is present in the first place (i.e., if $\gamma\rho$ is above a given threshold). A better understanding of evolution can then be achieved by looking at selection gradients (fig. 1*c*, 1*d*). These confirm that conditionality cannot rise from zero. Once $\rho$ crosses a threshold, however, it starts to increase by selection, which tends to favor even more conditionality, leading to an evolutionary runaway up to complete reciprocation ($\rho \approx 1$).

To complete the analysis, it is possible to evaluate the selection gradient on $\rho$ under the assumption that $\gamma$ is at its evolutionarily stable value (i.e., $S_\gamma = 0$) in order to determine the direction of selection on conditionality per se. The minimal value of $\rho$ that is necessary for selection to favor even more conditionality can then be plotted (fig. 2). This threshold increases with the cost of conditionality ($k$) and decreases with the amount of phenotypic variability ($\sigma^2$). Note that in practice the phenotypic variability of a trait is likely to depend quantitatively on the trait's mean value, which is not taken into account in the main text (but see "Evolutionary Dynamics with Phenotypic Variability" in the appendix).

*Interpretation.* In itself, cooperation, even in the presence of variability, does not select for conditionality, as there is no reason why it should be adaptive to cooperate more with more cooperative partners. Conditionality in any investment is beneficial if it allows individuals to adapt the amount invested to the expected return on investment. Hence, cooperating more with more cooperative partners is beneficial only if each unit of cooperation brings a larger return on investment with them, which is the case only if they already express their cooperation as a response to one's own cooperation. In this case, and only in this case, it may be worth cooperating more with more cooperative partners, because each unit of cooperation invested with them is more profitable. Consequently, (1) conditional cooperation cannot evolve from scratch, but (2) if a slight amount of conditionality arises for some other reason (see below), then a larger amount of conditional cooperation can be favored, as the cooperation of partners is now partially expressed as a response to one's own, which then increases the selective pressure favoring even more responsiveness, and so on. Consequently, even a very slight initial amount of conditionality can be enough to trigger runaway evolution toward full-fledged reciprocity ($\rho \approx 1$).

To understand how reciprocation can evolve from scratch, however, we need to explain how the initial degree of conditionality can arise. Various biological mechanisms can play a role here. They are presented in the "Discussion" section. Here I offer a formal illustration of one possibility.

*The Role of Synergy.* Conditionality can evolve from scratch (i.e., in the initial absence of conditionality) if it is beneficial in itself to cooperate more when one's partner also cooperates more, which can be formalized as a form of synergy. This mechanism can be introduced into the above model by slightly modifying the payoff function (eq. [3]) to be

$$P(h_., h_0) = ah_. + bh_0 - c(h_.)^2 + dh_.h_0. \qquad (9)$$

In this case, (i) both individuals may receive a benefit from

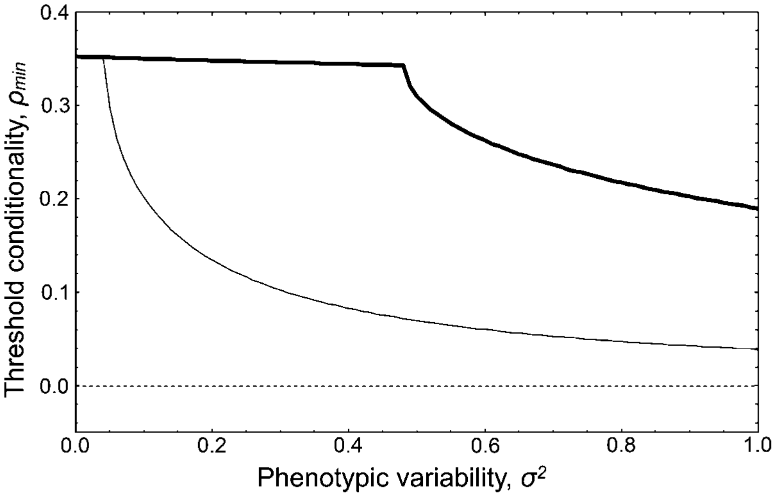


**Figure 2:** Selection for conditionality. Threshold value of conditionality above which selection starts favoring larger conditional abilities in the mechanistic model of the main text when $\gamma$ is evolutionarily stable, shown as a function of the amount of phenotypic variability $\sigma^2$ and for three values of the cost of conditionality: $k = 1$ (boldface curve), $k = 0.1$ (solid curve), and $k = 0$ (dashed curve). Other parameters are as in figure 1.

helping a little (because of common interests and/or genetic relatedness), and (ii) owing to the synergy term $dh.h_0$, they also receive an immediate benefit from adapting their level of helping to that of their partner. These two features are sufficient to trigger an initial rise in reciprocation, which can then lead to an evolutionary amplification (see appendix for mathematical details), which can also be observed by plotting the direction of selection as vector plots (fig. 3). Note that Akçay et al. (2009) have already observed the role of synergy in the evolution of reciprocity. In my analysis, however, we see more precisely that (i) synergy plays the role of a trigger needed only for the initial rise of some conditionality and (ii) the general requirement for the evolution of reciprocity is some responsiveness that preexists for reasons independent of reciprocity itself (and synergy is only one way to fulfill this requirement).

## Discussion

Reciprocity, in a broad sense, is characterized by individuals actively making it beneficial for their partners to cooperate by responding positively to cooperation (or responding negatively to defection). This makes it a particularly powerful generator of mutual benefits but also has the consequence that it entails a strong element of circularity, which renders its evolution problematic. For this reason, it is particularly important to understand how—through what steps and under what constraints—reciprocity can evolve. This may offer a hope of understanding the cause of the relative rarity of reciprocity in extant species and may help make sense of the instances in which it does occur.

### *Three Conditions for the Evolution of Reciprocity*

Reciprocity, even in a broad sense, always requires at least two essential ingredients: (i) the existence of a cooperative trait by which individuals provide benefits to others and (ii) an ability to express this trait conditionally. In this article, I have attempted to model the joint evolution of these two ingredients in the simple case of pairwise direct reciprocity (two individuals cooperate back and forth with each other; but see the appendix for a more general model). This analysis was performed under two different microscopic models of interactions, in which cooperation is either discrete or quantitative. The results are essentially similar in both cases. The models show that the evolutionary emergence of reciprocity entails three necessary conditions.

*Behavioral Variability.* The first condition is the least interesting and should be trivial (even though it is surprisingly neglected in many models; but see McNamara and Leimar 2010): there must be some variability in behavior. Conditional abilities in any domain can be selected for only provided there is some information worth responding to. It may be worth responding to some information about partners' cooperation provided there is some variation in this trait. For this reason, the models developed in this article assume the existence of some background variability in individuals' willingness to cooperate.

*Preexisting Cooperation.* The second condition is that some cooperation pre-evolves for reasons independent of reciprocity, for instance, because of the existence of genetic relatedness among partners or because cooperation immediately benefits the cooperator due to common interests (see also West et al. 2011, section 5.3). This constitutes a significant constraint regarding the situations in which reciprocal cooperation can rise. Whereas reciprocity allows cooperation to be adaptive even in the absence of kin selection or common interests, its initial rise requires one of these two mechanisms. However, this prerequisite is not sufficient for reciprocity to evolve.

*Preexisting Conditionality.* The third condition is highly constraining and is the least intuitive: selection can favor the ability to make cooperation conditional only if the cooperation expressed by others is already conditional. If partners simply cooperate more or less generously, independently of one's own level of cooperation, then there is no reason to adapt one's level of cooperation to theirs; that is, one should simply also cooperate unconditionally at the individually optimal level. If, on the other hand, partners respond more or less generously to one's cooperation, then each unit of investment spent with them is correspondingly more or less profitable, and it makes sense to adapt one's own level of cooperation to theirs. To put it simply, the conditionality of an individual's cooperation is made adaptive by the conditionality of others' cooperation.

This has two consequences, one negative and one positive. First, it constitutes a constraint on the situations in which reciprocal cooperation can evolve. Second, it generates a form of positive evolutionary feedback: an initially slight amount of preexisting conditionality can select for a stronger form of conditionality and hence for more cooperation, resulting in evolutionary amplification. Hence, reciprocal cooperation can be selected for, but it cannot be selected for from scratch. To trigger the process, some form of conditional cooperation must already be present, and it must be present for reasons that have nothing to do with its (future) role in triggering full-fledged reciprocity. The emergence of reciprocity is thus contingent on lucky initial conditions.

Note that even though most of the models developed in this article are models of so-called partner control (two individuals exchanging help back and forth; but see appendix, "An Even More General Model"), their results do also shed light on interactions involving partner choice (Noë et al. 1991). Indeed, partner control and partner

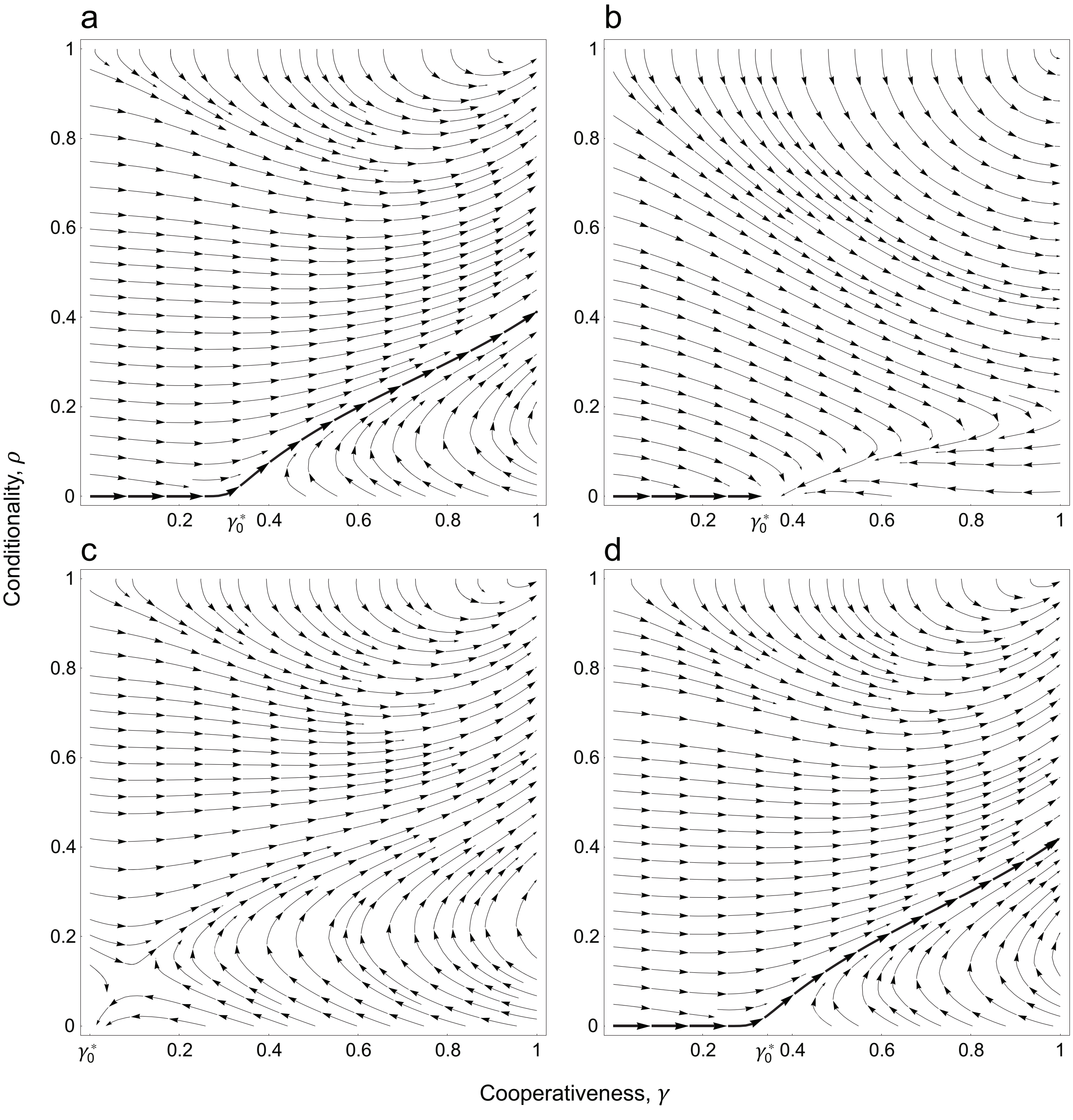


**Figure 3:** Direction of selection with synergy. Stream plots showing the direction of selection on cooperativeness $\gamma$ and conditionality $\rho$ in the mechanistic model of the main text in the presence of synergy ($d = 1$). As in figure 1, the vectors in boldface represent the evolutionary trajectory of a population seeded with $\gamma = 0$ and $\rho = 0$, and $\gamma_0^*$ is the evolutionarily stable cooperativeness in the absence of conditionality ($\rho = 0$). Constant parameters are $b = 10$, $c = 2$, $k = 0.1$, $d = 1$, and $R = 0$ (except in *d*). *a*, Phenotypic variability is present ($\sigma^2 = 0.5$), and cooperation has an automatic benefit ($a = 1$). *b*, Cooperation has an automatic benefit ($a = 1$), but phenotypic variability is absent ($\sigma^2 = 0$). *c*, Phenotypic variability is present ($\sigma^2 = 0.5$), but cooperation has no automatic benefit ($a = 0$). *d*, Phenotypic variability is present ($\sigma^2 = 0.5$), cooperation has no automatic benefit ($a = 0$), but genetic relatedness is positive ($R = 0.1$). In the presence of a sufficient amount of phenotypic variability and in the presence of either common interest ($a > 0$) or genetic relatedness ($R > 0$), unconditional cooperation first rises from 0 (because of *a* and/or *R*), then favoring conditionality (because of *d*), which then paves the way for the runaway increase of conditional cooperation.

choice differ only with regard to the precise adaptive reason for responding to one's partner. In partner control, one shall invest more with more responsive partners because one has other nonsocial activities that one can also invest in and that become comparatively less interesting when the partner is more responsive. In partner choice, on the other hand, one shall invest more with more responsive partners because one has other social activities (i.e., other partners) that become comparatively less interesting. Yet, the same principle applies in both cases. Cooperating more with more cooperative partners (including by choosing them) is adaptive provided their cooperation is expressed as a return on one's own, which implies that an initial form of responsiveness is already present for independent reasons.

In what follows, I discuss the major mechanisms that can play a triggering role as preexisting forms of responsiveness, and I show through examples that they do account for some instances of reciprocal cooperation observed in extant species.

### *Help to Help*

The first possibility is that cooperation initially provokes a positive response in others because it makes it easier, or simply possible, for them to cooperate. This can be understood through (partly imaginary) examples. Consider vampire bats exchanging blood meals or birds helping each other to mob predators and assume that some helping is initially favored owing to common interest or kin selection. In either case, helping has the peculiar property that it increases the probability of survival. Consequently, for a purely contingent (i.e., nonadaptive) reason, helping a partner increases the probability that she will be in a position to help later on because she is simply less likely to be dead (Eshel and Shaked 2001 and Kokko and Johnstone 2001 have considered this effect). The key point brought out by the models presented here is that such a contingent form of responsiveness can eventually select for an adaptive one. Because individuals respond to help by surviving, it is now worth giving more help to those who are themselves more helpful, because helping them is more beneficial. Individuals may thus evolve the ability to partly condition their helping to others' past helping. A genuinely conditional cooperation can then really evolve, but it requires the preexistence of a purely contingent form of conditionality, which plays the role of an evolutionary trigger.

More generally, any interaction in which helping makes it less costly, easier, or simply possible for others to help, that is, in which individuals help each other to help, entails such an initial trigger. This principle can apply when helping affects survival as in the above examples (Wilkinson 1988; Olendorf et al. 2004; Krams et al. 2008; and see Raihani and Bshary 2011 for a detailed discussion) but also when it affects the growth of one's partner, making future help more efficient (which could play a role in the establishment of mutualisms such as the plant-mycorrhiza interaction; Leimar and Connor 2003) or when it makes the other's helping less risky (as in predator inspection; Milinski 1987; see also Raihani and Bshary 2011 for discussion).

### *Synergy in Benefits*

A second possibility is that individuals benefit more from helping cooperative partners than others. A possible example is nuptial gifts, in which males offer resources to females in exchange for copulation. One possible scenario for the origin of nuptial gifts relies on the fact that males do benefit directly from helping females, but only if they have copulated with them (because this increases their probability of producing offspring who survive). This selects for a conditional ability in males, who should only give away their gift if copulation takes place. Again, following the general principle put forward in the models presented here, conditionality on one side selects for conditionality on the other. Conditionality in males selects for conditionality in females, who should now prefer to copulate with males who offer larger gifts, thereby increasing the selective pressure on the size of male gifts, and so on. Again, conditional cooperation preexists for reasons independent of reciprocity but may eventually allow the evolution of true conditionality on both sides.

### *Spatiotemporal Constraints*

In some cases, spatiotemporal constraints make it simpler, more practical, or even compulsory to cooperate in a symmetric fashion. In this case, cooperating with a partner does increase the probability that the partner will also cooperate, a simple form of responsiveness that can also play the role of an evolutionary trigger. An example in which this mechanism may have played a role is allogrooming (e.g., as observed in impalas by Hart and Hart 1992). The strong physical symmetry of grooming entails that grooming a partner (say, for an immediate benefit) makes grooming easier for the partner and thus more likely, which is a simple form of responsiveness. Eventually, this contingent form of responsiveness can then select for an adaptive one, as it creates a selective pressure to give priority to grooming those who are also in a grooming mood because the benefit of grooming is larger with them. Hence, the preexistence of some cooperation together with a spatial constraint can eventually yield the evolution of reciprocation.

### *Collective Action*

Another mechanism can force helping to be exchanged in a symmetric fashion. It occurs when the interaction is actually not an exchange but the production of a common good. In this case, by necessity, an individual can only receive a benefit from someone when also providing a benefit to them (these are the situations that scholars typically have in mind in the partner-choice literature). As in the case of spatiotemporal constraints, the preexisting conditionality is not an actual behavioral switch but rather the simple fact that interacting with a partner in one direction automatically forces the interaction to take place in the other direction as well.

A simple illustrative example is offered by the interaction between cleaner fishes and their clients (Bshary and Grutter 2006). The same single action (the cleaner eats the client's parasites) is a benefit that flows both from cleaner to client and from client to cleaner. In this case, clients have an immediate benefit in choosing to help good cleaners, which in turn makes it beneficial for cleaners to be more helpful.

More importantly, especially in the human case, this principle applies to any collective action in which several individuals produce a single benefit that then needs to be shared. In this case, as has been suggested several times (Sperber and Baumard 2012; Tomasello et al. 2012), individuals directly gain from conditionally helping the most helpful partners, because the collective benefit will be larger with them, which can lead to a rise in both cooperation and choosiness (see, e.g., McNamara et al. 2008).

Again, following the general logic of this article, even though responsiveness is initially a mere consequence of the ecology of collective action, it eventually triggers the evolution of genuine conditionality. First, it selects for a conditional ability to give priority to helping the most helpful partners. Second, this new conditionality selects for a further conditionality on the other side: in order to be chosen for cooperative ventures, individuals should now pay attention to the investment made by their partner in a collective action and then actively share the collective benefit accordingly so that their partner's eventual return on investment is satisfactory (Baumard et al. 2013). Hence, cooperation can eventually become conditional on all sides.

### *Helping as a Cue for Indirect Kin Recognition*

The last possibility is significantly different from the others. It is based on the fact that individuals can often recognize their kin by using indirectly the kin recognition abilities of others. Indeed, when some individuals have the ability to identify kin (e.g., parents recognize their offspring), then receiving help from them can be used as an indirect indication that they are close kin (human beings are known to recognize their younger siblings in such an indirect way; Lieberman et al. 2007). The interesting feature of such a strategy is that helping itself is a kin-recognition criterion and that individuals thus respond positively to helping for this very reason. Hence, secondarily, these mechanisms can be activated by nonkin to provoke the same beneficial response, potentially leading to reciprocal interactions. A likely instance of this mechanism is primate grooming (de Waal and Luttrell 1988; Barrett and Henzi 2001; Schino 2007; and see Raihani and Bshary 2011 for discussion). Grooming is known to provoke a relaxing physiological response that could have evolved initially in the context of kin relationships (because being groomed by someone is generally an indication that she is one's mother). This response eventually has positive effects for the groomer, which makes it adaptive to grooming even nonkin to provoke the same response (e.g., being tolerated at food sites by the groomee). Again, once in place, this can then select for even more conditionality, such as grooming specifically those who respond most positively to grooming, possibly leading to genuine reciprocal exchanges.

Note that this last mechanism is related to but significantly different from the proposal made by Axelrod and Hamilton (1981) that helping could be directly a kin-recognition criterion (helpers being more likely to be related to other helpers), which would also facilitate the evolution of reciprocity. However, Axelrod and Hamilton's (1981) proposal entails the maintenance of genetic polymorphism on helping itself (Rousset and Roze 2007; and see "Discussion on Strong Selection" in the appendix for a brief discussion). This constraining condition is relaxed if, as I suggest here, helping is a secondary consequence of other heuristics of kin recognition.

## Conclusion

The emergence of reciprocal cooperation requires that some cooperation and, furthermore, some conditional cooperation preexist for independent reasons, which has no general reason to be the case. Hence, reciprocal cooperation will not evolve each time constitutive cooperation is selected for. This helps make sense of a peculiar feature of the instances of reciprocal cooperation observed in extant species: in all of them, cooperation may have been conditional in the first place for reasons independent of reciprocity. This occurs, for instance, when helping in one direction makes it easier to help in the other direction (because of a spatial constraint) or when helping consists of a collective action in which a common good is produced and then shared. In itself, this result still leaves unexplained the extraordinary development of reciprocal cooperation in all sorts of contexts in the human species. Resolving this puzzle will be the object of a further study. It will likely require taking into account the cognitive mechanisms evolved in humans to manage reciprocity and their ability to function in a general manner, beyond the domains for which they have initially been selected.

## Acknowledgments

I thank C. El Mouden for very useful remarks on a previous version of this work and one anonymous reviewer for comments.

## Literature Cited

Akçay, E., J. Van Cleve, M. W. Feldman, and J. Roughgarden. 2009. A theory for the evolution of other-regard integrating proximate and ultimate perspectives. Proceedings of the National Academy of Sciences of the USA 106:19061–19066.

André, J.-B. 2010. The evolution of reciprocity: social types or social incentives? American Naturalist 175:197–210.

———. 2014. Mechanistic constraints and the unlikely evolution of reciprocal cooperation. Journal of Evolutionary Biology 27:784–795.

André, J.-B., and T. Day. 2007. Perfect reciprocity is the only evolutionarily stable strategy in the continuous iterated Prisoner's Dilemma. Journal of Theoretical Biology 247:11–22.

Axelrod, R., and W. D. Hamilton. 1981. The evolution of cooperation. Science 211:1390–1396.

Barrett, L., and S. P. Henzi. 2001. The utility of grooming in baboon troops. Pages 119–145 *in* R. Noë, J. A. R. A. M. Van Hooff, and P. Hammerstein, eds. Economics in nature. Cambridge University Press, Cambridge.

Baumard, N., J.-B. André, and D. Sperber. 2013. A mutualistic approach to morality: the evolution of fairness by partner choice. Behavioral and Brain Sciences 36:59–122.

Bergmuller, R., R. A. Johnstone, A. F. Russell, R. Bshary, and R. Bergm. 2007. Integrating cooperative breeding into theoretical concepts of cooperation. Behavioural Processes 76:61–72.

Brauchli, K., T. Killingback, and M. Doebeli. 1999. Evolution of cooperation in spatially structured populations. Journal of Theoretical Biology 200:405–417.

Bshary, R., and A. S. Grutter. 2006. Image scoring and cooperation in a cleaner fish mutualism. Nature 441:975–978.

Clutton-Brock, T. 2009. Cooperation between non-kin in animal societies. Nature 462:51–57.

Connor, R. C. 1986. Pseudo-reciprocity: investing in mutualism. Animal Behaviour 34:1562–1566.

———. 1995*a*. Altruism among non-relatives: alternatives to the Prisoner's Dilemma. Trends in Ecology and Evolution 1:84–86.

———. 1995*b*. The benefits of mutualism: a conceptual framework. Biological Reviews 70:427–457.

de Waal, F. B., and L. M. Luttrell. 1988. Mechanisms of social reciprocity in three primate species: symmetrical relationship characteristics or cognition? Ethology and Sociobiology 9:101–118.

Eshel, I., and A. Shaked. 2001. Partnership. Journal of Theoretical Biology 208:457–474.

Hammerstein, P. 2003. Why is reciprocity so rare in social animals? a protestant appeal. Pages 83–94 *in* P. Hammerstein, ed. Genetic and cultural evolution of cooperation. MIT Press, Cambridge, MA.

Hart, B. L., and L. A. Hart. 1992. Reciprocal allogrooming in impala, *Aepyceros melampus*. Animal Behaviour 44:1073–1083.

Hauert, C., and H. G. Schuster. 1997. Effects of increasing the number of players and memory size in the iterated Prisoner's Dilemma: a numerical approach. Proceedings of the Royal Society B: Biological Sciences 264:513–519.

Hauert, C., and O. Stenull. 2002. Simple adaptive strategy wins the Prisoner's Dilemma. Journal of Theoretical Biology 218:261–272.

Imhof, L. A., D. Fudenberg, and M. A. Nowak. 2005. Evolutionary cycles of cooperation and defection. Proceedings of the National Academy of Sciences of the USA 102:10797–10800.

Imhof, L. A., and M. A. Nowak. 2010. Stochastic evolutionary dynamics of direct reciprocity. Proceedings of the Royal Society B: Biological Sciences 277:463–468.

Kandori, M., G. Mailath, and R. Rob. 1993. Learning, mutation, and long run equilibria in games. Econometrica 61:29–56.

Kokko, H., and R. A. Johnstone. 2001. The evolution of cooperative breeding through group augmentation. Proceedings of the Royal Society B: Biological Sciences 268:187–196.

Krams, I., T. Krama, K. Igaune, and R. Mänd. 2008. Experimental evidence of reciprocal altruism in the pied flycatcher. Behavioral Ecology and Sociobiology 62:599–605.

Lehmann, L., and L. Keller. 2006. The evolution of cooperation and altruism: a general framework and a classification of models. Journal of Evolutionary Biology 19:1365–1376.

Leimar, O. 1997. Reciprocity and communication of partner quality. Proceedings of the Royal Society B: Biological Sciences 264:1209–1215.

Leimar, O., and R. Connor. 2003. By-product benefits, reciprocity, and pseudoreciprocity in mutualism. Pages 203–222 *in* P. Hammerstein, ed. Genetic and cultural evolution of cooperation. MIT Press, Cambridge, MA.

———. 2010. Cooperation for direct fitness benefits. Philosophical Transactions of the Royal Society B: Biological Sciences 365:2619–2626.

Lieberman, D., J. Tooby, and L. Cosmides. 2007. The architecture of human kin detection. Nature 445:727–731.

McNamara, J. M., Z. Barta, L. Fromhage, and A. I. Houston. 2008. The coevolution of choosiness and cooperation. Nature 451:189–192.

McNamara, J. M., Z. Barta, and A. I. Houston. 2004. Variation in behaviour promotes cooperation in the Prisoner's Dilemma game. Nature 428:745–748.

McNamara, J. M., and O. Leimar. 2010. Variation and the response to variation as a basis for successful cooperation. Philosophical Transactions of the Royal Society B: Biological Sciences 365:2627–2633.

Milinski, M. 1987. Tit for tat in sticklebacks and the evolution of cooperation. Nature 325:433–435.

Noë, R., J. A. Van Hooff, and P. Hammerstein. 2001. Economics in nature: social dilemmas, mate choice and biological markets. Cambridge University Press, Cambridge.

Noë, R., C. P. Vanschaik, and J. A. R. A. M. Van Hooff. 1991. The market effect: an explanation for pay-off asymmetries among collaborating animals. Ethology 87:97–118.

Nowak, M. A., A. Sasaki, C. Taylor, and D. Fudenberg. 2004. Emergence of cooperation and evolutionary stability in finite populations. Nature 428:646–650.

Nowak, M. A., and K. Sigmund. 1992. Tit for tat in heterogeneous populations. Nature 355:250–253.

———. 1993. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. Nature 364:56–58.

———. 1994. The alternating Prisoner's Dilemma. Journal of Theoretical Biology 168:219–226.

———. 1995. Invasion dynamics of the finitely repeated Prisoner's Dilemma. Games and Economic Behavior 11:364–390.

Olendorf, R., T. Getty, and K. Scribner. 2004. Cooperative nest defence in red-winged blackbirds: reciprocal altruism, kinship or by-product mutualism? Proceedings of the Royal Society B: Biological Sciences 271:177–182.

Orr, H. A. 2005. The genetic theory of adaptation: a brief history. Nature Reviews Genetics 6:119–127.

Raihani, N. J., and R. Bshary. 2011. Resolving the iterated Prisoner's Dilemma: theory and reality. Journal of Evolutionary Biology 24: 1628–1639.

Roberts, G., and T. N. Sherratt. 1998. Development of cooperative relationships through increasing investment. Nature 394:175–179.

Rousset, F. 2004. Genetic structure and selection in sudivided populations. Princeton University Press, Princeton, NJ.

Rousset, F., and D. Roze. 2007. Constraints on the origin and maintenance of genetic kin recognition. Evolution 61:2320–2330.

Schino, G. 2007. Grooming and agonistic support: a meta-analysis of primate reciprocal altruism. Behavioral Ecology 18:115–120.

Sperber, D. A. N., and N. Baumard. 2012. Moral reputation: an evolutionary and cognitive perspective. Mind and Language 27: 495–518.

Stevens, J. R., F. A. Cushman, and M. D. Hauser. 2005. Evolving the psychological mechanisms for cooperation. Annual Review of Ecology, Evolution, and Systematics 36:499–518.

Stevens, J. R., and M. D. Hauser. 2004. Why be nice? psychological constraints on the evolution of cooperation. Trends in Cognitive Sciences 8:60–65.

Tomasello, M., A. P. Melis, C. Tennie, E. Wyman, and E. Herrmann. 2012. Two key steps in the evolution of human cooperation. Current Anthropology 53:673–692.

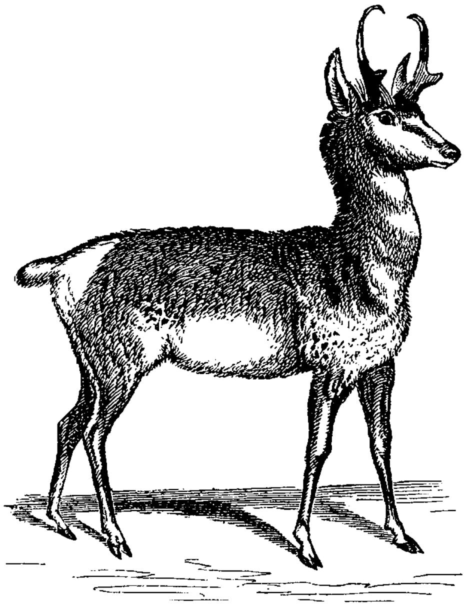
Trivers, R. L. 1971. The evolution of reciprocal altruism. Quarterly Review of Biology 46:35–57.

West, S. A., C. El Mouden, and A. Gardner. 2011. Sixteen misconceptions about the evolution of cooperation in humans. Evolution and Human Behavior 32:231–262.

West, S. A., A. S. Griffin, and A. Gardner. 2007. Evolutionary explanations for cooperation. Current Biology 17:R661–R672.

Wilkinson, G. S. 1988. Reciprocal altruism in bats and other mammals. Ethology and Sociobiology 9:85–100.

Associate Editor: Sean H. Rice
Editor: Susan Kalisz



"The antelope fawns are born in the spring, and when six months old the horns first begin to develop. They continue to grow until the next October or November (that is, until the animal is eighteen months old), when the first pair of horns are shed; by this is meant the outer shell." From "The Prong-Horn Antelope" by W. J. Hays (*The American Naturalist*, 1868, 2:131–133).