

# THE EVOLUTION OF FAIRNESS IN A BIOLOGICAL MARKET

Jean-Baptiste André<sup>1,2</sup> and Nicolas Baumard<sup>3,4,5</sup>

<sup>1</sup>CNRS, Ecole Normale Supérieure, UMR 7625, "Écologie and Évolution", 46 rue d'Ulm, 75005 Paris, France

<sup>2</sup>E-mail: jeanbaptisteandre@gmail.com

<sup>3</sup>Institute of Cognitive and Evolutionary Anthropology, University of Oxford, 64 Banbury Road, Oxford OX2 6PN, United Kingdom

<sup>4</sup>E-mail: nbaumard@gmail.com

<sup>5</sup>Institut Jean-Nicod, UMR 8129, CNRS, École Normale Supérieure, École des Hautes Études en Sciences Sociales, 29, rue d'Ulm, 75005, Paris, France

Received May 20, 2010

Accepted December 4, 2010

Human beings universally express a concern for the fairness of social interactions, and it remains an open question that which ultimate factors led to the evolution of this preference. Here, we present a model accounting for the evolution of fairness on the basis of individual selection alone. We consider a simple social interaction based on the Dictator Game. Two individuals, a "proposer" and a "responder," have an opportunity to split a resource. When they have no choice but to interact together, the most powerful (here the proposer) reaps all the profits and fairness cannot evolve. Partner choice is the key lever to overcome this difficulty. Rather than just two individuals, we consider a population composed of two classes of individuals (either proposers or responders), and we allow the responders to choose their partner. In such a "biological market," fairness evolves as an "equilibrium price," resulting from an ecological equivalent of the law of supply and demand. If a class is disadvantaged by the chosen resource partition (i.e., if it frequently receives less than half of the resource), it is outcompeted by the other one, and automatically becomes rarer. This rarity grants it an advantage on the market, which yields in turn to the evolution of a more favorable partition. Splitting the resource into two identical halves, or more generally in a way that equalizes the payoffs of the two classes, is then the only evolutionarily stable outcome. Beyond human fairness, this mechanism also opens up new ways of explaining the distribution of benefits in many mutualistic interactions.

**KEY WORDS:** Evolution of cooperation, game theory, moral psychology, mutualism, partner-choice, ultimatum game.

The evolution of cooperation among nonrelated partners, especially in humans, has been the object of intense modeling and debates in recent decades. The vast majority of work in this area aims to understand how natural selection may lead to the emergence and maintenance of cooperation (Axelrod 1981; Nowak and Sigmund 1992, 1998; Leimar and Hammerstein 2001; Lehmann and Keller 2006; Nowak 2006). Yet, although this body of work has been central to our understanding of the existence of cooperation, it is also essential to understand how natural selection has shaped the "way" we cooperate, and in particular the way

we share its benefits. Indeed, whereas the surplus generated by a cooperative interaction can in principle be distributed in infinitely many ways, human beings systematically prefer certain, so-called "fair," distributions (e.g., half-and-half splits in symmetric interactions). The aim of the present article is to explore how natural selection has shaped these preferences.

## ASYMMETRIC NEGOTIATION

With this aim, taking a step further relative to most models on cooperation, we take for granted the occurrence of a cooperative

interaction between two players, generating a surplus of constant size, and we seek to understand how partners distribute this surplus (under the assumption that it is perfectly divisible). As a particularly simple negotiation mechanism, we consider a strategically asymmetric interaction (i.e., asymmetric with regard to negotiation power) in which one player, called the “proposer,” has the power to definitely commit to a division of the resource, with no chance of changing his mind afterward, and the other, called the “responder,” has no better option than to accept it (see Schelling 1960). The paradigmatic examples of such negotiations are the well-known dictator and ultimatum games (DG and UG; Güth et al. 1982; Camerer 2003). In both games, the proposer commits to an offer that the responder can only accept (in the DG) or can either accept or entirely refuse (in the UG).

The evolution of each partner’s strategy in such negotiations is straightforward to understand (see also Nowak et al. 2000). On one hand, whatever offer a responder’s partner makes, she gains more resources if she accepts it than if she rejects it. Therefore, in all cases, natural selection favors indiscriminate responders taking whatever resources are made available to them. On the other hand, and as a result, selection favors stingy proposers offering the minimal possible amount. The resource division at evolutionary equilibrium is thus maximally “unfair”: the empowered individual (the proposer) keeps virtually all the resource.

Admittedly, in less asymmetric, more complex, and often more realistic, negotiation processes, in which each partner can successively reject the other’s offer and make his own afterward, the equilibrium split is less asymmetric (Stahl 1972); and when both players have the exact same negotiation power the equilibrium split is just fair, (Rubinstein 1982). But what we aim to explain here is the fact that human beings express concerns for fairness even when negotiation power is unequal. In social life, taking advantage of a local strategic advantage is often viewed as a paradigmatic instance of unfairness (e.g., stealing in a shop because the shopkeeper is old and weak). In brief, the distribution of resources among humans is typically not the outcome of a mere power struggle, and this is what we aim to understand. Highly asymmetric interactions are thus particularly interesting for us as a test case.

### **REPUTATION WITHOUT PARTNER CHOICE: THE REVERSAL OF NEGOTIATION POWER**

In a paper in *Science*, Nowak et al. (2000) modified the ultimatum game by introducing a mechanism akin to reputation, and this paper is often considered as theoretical evidence that “reputation” in general can lead to fairness in this game. Yet the model presented actually offers no such evidence, and it is important to understand why.

Nowak et al. (2000) consider a model in which responders have a reputation consisting in the resource division they are ready

to accept. Some responders are known for being very demanding (they reject even high offers), whereas others are known for being very compliant. Pairs of partners are formed randomly, and the proposer is informed of his responder’s “reputation” and makes an offer. The offer is then accepted or rejected in accordance with the responder’s reputation.

Via their reputation, responders now have the power to commit to a given offer (e.g., “any offer below 90% will be rejected”; “reputation” is hence just another term for “commitment”), and proposers have no other choice but to comply. In other words, responder’s reputation, far from removing the strategic asymmetry of the game, completely reverses it. Because they are the first to speak, responders now effectively play the role of proposers, and vice versa. The evolutionary outcome is thus also reversed. Natural selection favors compliant proposers offering whatever resources are requested from them, and demanding responders asking for the maximal possible amount. The resource division at evolutionary equilibrium is still maximally unfair: the responder keeps all.

Fairness nevertheless emerges in Nowak et al. (2000) as a consequence of an assumption. Nowak et al. (2000) a priori restrict the parameter space, such that individuals cannot obtain more resources when they are in the role of responders than when they are proposers, an assumption chosen precisely to prevent responders from keeping more than half of the resource at equilibrium, which would be a “perhaps unrealistic complication” (Nowak et al. 2000, note 14). As a result, when responders are advantaged by the power struggle because they are the first to commit, natural selection leads them to obtain the largest possible amount of resources within the limits authorized by the model. In this case, they take just half of the resource.

### **PARTNER CHOICE**

When two partners have no other choice but to interact with each other, the strategic power struggle is determinant. If the strategic relationship is completely asymmetric (as in the DG, the UG, or the “reversed UG” of Nowak et al. 2000), fairness cannot evolve, unless it is made compulsory by constraints on the parameter space.

In this article, we aim to show that fairness can nevertheless evolve in the division of a resource, even when the interaction is strategically asymmetric, provided responders have the possibility to “choose” their preferred partner, a mechanism that has been shown empirically to play an important role in human cooperation (e.g., Barclay 2004; Barclay and Willer 2007; Chiang 2010). When responders can choose their partner, social life resembles a “biological market” where proposers compete to be chosen (Noë and Hammerstein 1994). We develop a series of models showing how fairness can evolve in such a market. We show that, far from being a trivial outcome of partner choice, the evolution of fairness

is the consequence of a complex interaction between resource division and the demography of each trading class (responders and proposers).

## The Models

The basic model we consider is based upon the dictator game, except that we consider a population of proposers interacting with the members of a population of responders that can choose among them. Each proposer is genetically characterized by his offer  $p$ . In each round, the proposers present in the population make their offer in public. As in the traditional dictator game, they are committed to this offer. Individual responders are then taken one by one in a random order, and asked to choose a proposer among those that are still unpaired at this time (if any). Once a proposer has been chosen, he is removed from the pool of available proposers (i.e., each individual is paired at most with one partner). For responders, the best strategy consists in picking the most generous proposer available, and accepting his offer. We hence assume that all responders follow this strategy (but see Models 3 and 5) and we focus on the evolution of proposers' offers. The advantage of small offers is to keep more of the resource for oneself. The advantage of large offers is to attract partners. These two opposite selective forces drive the evolution of  $p$ .

Note that the very same model can also be described in terms of reputation (see e.g., Nowak et al. 2000). Proposers' actual offers are publicly observed so that when a proposer is genetically characterized by an offer  $p$  he also has the public "reputation" of offering  $p$ . Responders then choose their partner on the basis of his reputation.

All models are deterministic resident-mutant analyses (Maynard Smith and Price 1973). We consider a population of infinite size fixed with a resident strategy and challenge it by introducing mutants, until we find a resident that no mutant can invade. This approach is valid under the assumption that mutants rarely occur. Resident populations are challenged by mutants one at a time, and novel mutants always appear after the fixation of a previous mutant is complete (Geritz et al. 1998).

However, because frequency-dependent effects are to be expected, we do not restrict our analysis to describing the fate of infinitely rare mutants. We describe the effect of selection for mutants in any frequency  $x \in [0, 1]$ . We check (1) whether mutants that are favored when rare reach complete fixation or remain at an intermediate frequency, and (2) whether mutants that are counter-selected when rare can nevertheless invade if they are initially introduced at a larger frequency.

We develop five distinct models based on these principles (Models 1–5). Model 1 is fully presented in the article as an illustration of the general principles underlying our analyses; all other models are described in detail in the Supporting information (SI).

### MODEL 1: BIOLOGICAL MARKETS DO NOT ALWAYS PROMOTE FAIRNESS

In Model 1, we consider a population in which the density of individuals in each trading class— $n_p$  proposers and  $n_r$  responders—is constant (Noë and Hammerstein 1994). If  $n_r > n_p$ , some responders will end up with no partner, and conversely if  $n_p > n_r$ . We show that fairness cannot evolve in this case.

We consider a population with a density  $n_p$  of proposers, a fraction  $1 - x$  of which are resident individuals offering  $p$ , and a fraction  $x$  are mutants offering  $p' = p + \delta$ . We aim to compare mutants' and residents' payoffs. Calling  $r$  the relative density of responders ( $r = n_r/n_p$ ), we delineate two cases.

#### Proposers are limiting ( $r \geq 1$ )

In this case, every proposer ends up paired with a responder. Residents' and mutants' payoffs are  $P = 1 - p$  and  $P' = 1 - p'$ , respectively. Mutants increase in frequency if and only if  $p' < p$ . Evolution leads to the ultimate fixation of the smallest possible offer,  $p = 0$ .

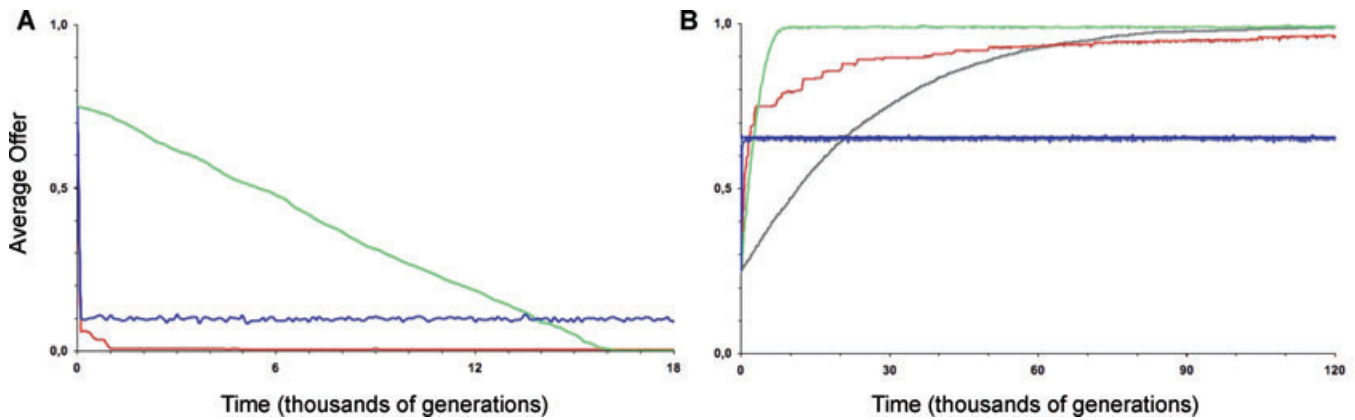
#### Responders are strictly limiting ( $r < 1$ )

In this case, some proposers (the least generous) end up with no partner. We then delineate two subcases, depending on the nature of mutants.

(1) If mutants are more generous than residents ( $p' > p$ ), mutants are preferred. When the density of generous mutants is strictly lower than the density of responders (i.e., when  $x < r$ ), every mutant finds a partner, and some residents do. Residents' and mutants' payoffs are  $P = (1 - p)(r - x)/(1 - x)$  and  $P' = 1 - p'$ , respectively. Generous mutants increase in frequency if and only if  $\delta < (1 - p)(1 - r)/(1 - x)$ . There always exist generous mutants satisfying this condition (recall that  $r < 1$ ). When the density of mutants becomes larger than that of responders (i.e., when  $x \geq r$ ), only mutants end up with a partner (but not necessarily all of them). Residents' and mutants' payoffs are  $P = 0$  and  $P' = (r/x)(1 - p')$ , respectively. Mutants increase in frequency until complete fixation.

(2) The very same analysis can be applied to the inverse situation where mutants are less generous than residents ( $p' < p$ ). If stingy mutants are rare (precisely if  $x \leq 1 - r$ ), they are counter-selected and decrease in frequency until complete extinction. If stingy mutants are frequent (precisely if  $x > 1 - r$ ), they increase in frequency if and only if their difference with the resident is  $\delta < (1 - p)(r - 1)/(r - 1 + x)$ . In this case they reach complete fixation.

Overall, when mutants are infinitely rare initially ( $x = 0$ ), only generous mutants can increase in frequency when  $n_p > n_r$ , and only stingy mutants can increase in frequency when  $n_p \leq n_r$ . Evolution leads to the fixation of the largest possible offer ( $p = 1$ )



**Figure 1.** Evolution of the average offer in simulations, when the densities of proposers and responders are held constant. Each curve is a single simulation run. In black, mutations are rare (the mutation probability is  $10^{-4}$  per generation) and of weak effect (mutations have a normally distributed effect with standard deviation  $10^{-3}$ ). In red, mutations are rare and 90% have a strong effect (mutants' offer is taken from a uniform distribution between 0 and 1). In green, mutations are frequent (the mutation probability is  $10^{-1}$  per generation) and of weak effect. In blue, mutations are frequent and 90% have a strong effect. Population size is  $10^4$  and nonsocial payoff  $10^{-5}$ . (A) The population contains 45% of proposers. The offer is initially 0.75. Evolution leads to the reduction of the average offer down to 0 in all cases. In blue, the deviation from 0 is due to mutation–selection balance. When mutations are rare and of weak effect, evolution down to 0 is very slow (more than  $10^6$  generations). (B) The population contains 55% of proposers. The offer is initially 0.25. Evolution leads to the increase of the average offer up to 1 in all cases except when mutations are both frequent and of strong effect (blue) because very stingy mutants can increase in frequency when sufficiently frequent (see SI).

when  $n_p > n_r$ , and the lowest possible offer ( $p = 0$ ) when  $n_p \leq n_r$ . The same occurs when mutants are very weakly different from residents ( $\delta \approx 0$ ), whatever their initial frequency. The outcome is more complex when (1) mutants' frequency is not infinitely small and (2) mutants' effect is not very weak. In this case, evolution toward  $p = 0$  when  $n_p \leq n_r$  is still guaranteed, but evolution toward  $p = 1$  when  $n_p > n_r$  is not. Stochastic individual-based simulations (see SI for details on the simulation process) confirm the analytical results (Fig. 1). When the mutation rate and/or the effect of mutation are low, evolution leads toward  $p = 0$  or  $p = 1$  (depending on the respective values of  $n_p$  and  $n_r$ ). When the mutation rate and the effect of mutations are both large, evolution leads toward  $p = 0$  when  $n_p \leq n_r$  but not to  $p = 1$  when  $n_p > n_r$ .

In real-life settings, mutants are expected to be introduced in very small frequencies. In this case, the least frequent of the two trading classes always monopolizes the benefits of the interaction. The strategic power struggle does vanish when proposers are competing for access to partners, only to be replaced by a “market power struggle,” which is just as unfair.

## MODEL 2: FAIRNESS EVOLVES WHEN SUPPLY AND DEMAND DEPEND ON PRICES

In microeconomic theory, prices reach equilibrium because they affect supply and demand. When prices are too high, demand shrinks and supply rises, which pushes prices down. Conversely, when prices are too low, demand rises and supply shrinks, which pushes prices up. There can be no equilibrium without this feedback.

In our model, “prices” correspond to offers ( $p$ ), “supply” to the density of proposers, and “demand” to the density of responders. So far, prices affect neither supply nor demand, because the densities of proposers and responders are supposed to be constant (and because, irrespective of the offers made, every responder wants to interact with exactly one proposer, and vice versa). As a result, there is no equilibrium. Any market asymmetry (one trading class is rarer than the other) inevitably leads to a maximally unfair distribution of the resource (if mutants are rare and/or weakly different from residents), because the market effect is consistently pushing in the same direction.

Yet in reality, prices do feed back on supply and demand, even in biology. This can occur in two sorts of ways. Individuals can change the quantity of social interaction they entertain in response to the “price” (this occurs for instance in Johnstone and Bshary 2008). But, in biology, the simplest and most general form that such feedback may take is demography: prices affect the density of individuals of each trading class (this is thus an instance of a demographic feedback in a game, see e.g., Hauert et al. 2006 for another). When division of the resource is biased in favor of one class, the individuals of this class obtain a larger payoff than others. The frequency of this class is then likely to increase. It may occur for two reasons. (1) Trading class is heritable from parent to child. If a class has a larger average payoff than the other, its frequency increases owing to natural selection. This occurs for instance if the trading classes are two different species engaged in some form of mutualism. (2) An individual's trading class is the product of a decision based on various sources of information

about average payoffs. The class that obtains the largest payoff tends to be preferred, and becomes more frequent. This occurs in the case of human behavior, when two different social roles interact to produce a common good. In either case, the change in class frequency compensates the inequality of payoffs. When a trading class becomes more frequent than the other, its average payoff decreases because individuals of this class have more and more difficulty finding a partner. The frequency of each class thus stabilizes precisely when the average payoffs to each are equal. This is the “ecological equilibrium” of class frequency. In our model, it is straightforward to show that, for a given average offer  $p$ , the frequency of responders at ecological equilibrium is  $p$ , and that of proposers  $1 - p$ .

Model 2 aims to analyze the effect of this ecological feedback on the evolution of offers. To do so, we assume that the ecological equilibrium of class frequency is reached very rapidly relative to the rate of evolution of offers (i.e., ecology is faster than evolution). This assumption is adapted to our primary center of interest: intraspecific mutualism in humans, where individuals plastically choose the role they wish to play. Model 2 thus follows the same principles as Model 1, except that the relative density of responders is assumed to be always equal to its ecological equilibrium value. In complementary simulations (see SI), we undertake the opposite assumption, where both an individual’s trading class and his offer are genetically encoded and evolve through natural selection; this opposite assumption is adapted for instance to the case of a mutualistic interaction between two different species. Details of the analyses are presented in SI, and simulation results in Figure S1. Here, we present the four major results stemming from the analysis of Model 2.

(1) In a resident population of stingy proposers ( $p < 0.5$ ), more generous proposers are able to invade, whereas less-generous ones can never increase in frequency. (2) In a population of generous proposers ( $p > 0.5$ ), less-generous proposers are able to invade, whereas more generous ones can never increase in frequency. (3) In a population of exactly fair proposers no other strategy can increase in frequency. (4) Even though stable mixtures of generous and stingy proposers do exist, fair proposers ( $p = 0.5$ ) can always invade them. In consequence, fairness does evolve from any initial state, and is evolutionarily stable. Interestingly, in contrast with Model 1, this result is independent of the frequency and effects of mutants. Simulations are in line with analytical results (Fig. S1).

This outcome is relatively easy to grasp intuitively. Consider first a resident population of selfish proposers. At ecological equilibrium, responders are limiting because they are disadvantaged by the resource division, and the market thus favors more generous proposers (see Model 1). This goes on until the resource is split exactly in two identical halves, and the frequency of each trading class becomes identical. Conversely, in a resident popula-

tion of generous proposers, proposers are limiting, and the market favors less-generous offers until the resource is also split exactly in two halves. Fairness is the only equilibrium.

Note that Nowak et al. (2000) also implicitly consider a situation in which proposers and responders are in equal density (because individuals are randomly assigned a role with “equal” probability). However, it would be a mistake to believe that this explains the evolution of fairness in their model (as explained in the introduction, it is explained by a restriction of parameter space). In the absence of partner choice, only local strategic issues matter, and the relative density of proposers and responders has no impact on the evolutionary outcome. The equalization of role density leads to fairness in our model only via partner choice.

In what follows, we describe two complementary models aimed at testing the robustness and generality of our result. For the sake of simplicity, in contrast with Model 2, these models include an assumption of weak selection. We assume that the effects of mutations are infinitesimally small, and we therefore only consider the first-order effects of offers on payoffs.

### MODEL 3: IMPERFECT RESPONDERS

Models 1 and 2 consider “perfect” responders able to discriminate infinitely precisely among all offers. Under this assumption, only slightly increasing one’s offer drastically increases one’s probability of finding a partner. This generates a strong incentive in favor of generosity. In reality, however, responders are probably not able to discriminate so precisely between offers. Furthermore, they may not be informed of every offer available in the population. The benefit of generosity could be reduced in either of these cases.

In Model 3, we aim to verify that our results still hold under this more realistic assumption. We build a quantitative model of choice, in which a given proposer’s probability of finding a partner is a continuous function of the difference between his own offer and the average offer in the population. The analysis shows that the essential mechanism promoting fairness still works (see SI for details). Fairness evolves as long as a proposer’s probability of being chosen as a partner increases sufficiently strongly with his offer, and even if responders discriminate between offers with limited precision.

### MODEL 4: FAIRNESS IN ASYMMETRIC INTERACTIONS

In the above sections, we considered an interaction in which both partners benefit identically from each unit of a resource. We showed that the evolutionarily stable resource division is then symmetric (half-and-half). In many real-life cases, however, partners’ payoffs are asymmetric. The evolutionarily stable resource division might then also be asymmetric. To study this more general case, we developed a further model.

Consider an interaction between a proposer and a responder, in which the responder gets a fraction  $p$  of the resource. The proposer receives a payoff of  $P_p = b_p(1 - p) - c_p$ , and the responder one of  $P_r = b_r p - c_r$ , where  $b_i$  is the marginal benefit of each resource unit and  $c_i$  the cost of the social interaction for individuals of type  $i$ .

The analysis of Model 4 (see details in SI) shows that the half-and-half split obtained in previous models is a special case of a more general outcome. In both symmetric and asymmetric interactions, evolution leads to resource divisions that “equalize” the total payoff of individuals (i.e., such that  $P_p = P_r$ ). Mathematically, the evolutionarily stable offer is thus  $p^* = (b_p + c_r - c_p)/(b_p + b_r)$ . In the special case in which payoffs are symmetric ( $b_p = b_r$  and  $c_p = c_r$ ) this corresponds to an equal split, but this need not be the case in general. For instance, if responders have to invest more energy to produce the common resource ( $c_r > c_p$ ), this initial investment ends up being fully compensated by an asymmetric division of the resource in their favor (i.e.,  $p^* > 0.5$ ), and vice versa.

Interestingly, the asymmetric case helps to highlight the divergences between the outcome of evolution in a biological market and social optima. Assume for instance that proposers gain more for each unit of resource they receive ( $b_p > b_r$ ). The sum of individual payoff would be maximized if proposers received the entire resource ( $p = 0$ ); in economics, this is defined as the “social optimum.” But evolution in a biological market does not lead to such a biased distribution. On the contrary, if proposers gain more from each unit of a resource, they receive fewer resources at equilibrium ( $p^* > 0.5$ ), and this is socially suboptimal (i.e., it does not maximize the total payoff).

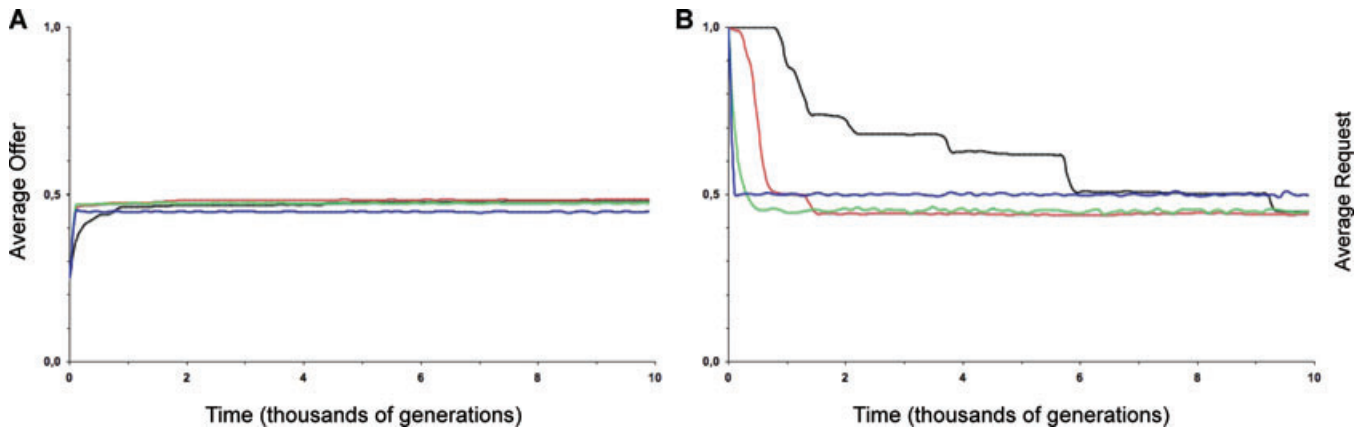
#### MODEL 5: FAIRNESS ON THE SIDE OF RESPONDERS

In Models 2–4, proposers evolve a preference for fairness owing to a competition for the access to responders. Responders, on the other hand, merely pick the best offer available and do not express concern for fairness. Selection may nevertheless promote a preference for fairness on the side of responders if there is also a competition “among responders” for the access to proposers. Such a competition is likely to take place in real-life settings. When a responder meets a proposer, she can either accept him as a partner, or reject his offer and wait until she (hopefully) comes across a more generous proposer in the future. In the latter case, she runs the risk of either ending up without any partner because they have all been paired in the meantime, or of being forced to eventually accept a lower offer than the one she initially rejected. In other words, competition among responders creates an “urgency” effect. Responders might be better off not assessing all available offers, but instead setting a minimum request: a resource division that they are ready to accept without comparing further offers.

Model 5 aims to capture this phenomenon. Consider again a population with  $n_p$  proposers and  $n_r$  responders, and consider a pairing procedure with two stages. In the “first stage,” pairs are formed randomly, with one proposer and one responder (some individuals remain unpaired if  $n_p \neq n_r$ ). In each pair, the proposer makes an offer,  $p$ , that the responder accepts or rejects (here, the rejection of offers is hence explicitly considered). If the offer is accepted, it is implemented; otherwise the pair of players moves on to the second pairing stage. The “second stage” involves the individuals that have not been successfully paired in the first, and is strictly equivalent to the single stage of previous models: each responder chooses the most generous proposer among those available. Proposers are characterized by their offer ( $p$ ), and responders by the minimum offer they are ready to accept in the first stage ( $q$ ). To generate a selective pressure on the responder’s side, we need to introduce some background variability among offers, and we do so by assuming that proposers tremble and deviate slightly from their expected offer (see also Noë and Hammerstein, 1994).

Ultimately, we aim to study the joint evolution of  $p$  and  $q$  when the relative density of responders,  $r$ , varies dynamically with payoffs. To study this complex process, we used stochastic individual-based simulations, under the assumption that individuals’ offer, request, and trading class are all genetically encoded (see SI for details). We also developed simple mathematical arguments (detailed in SI) to describe (1) the evolution of  $q$  when  $p$  and  $r$  are held constant, and (2) the evolution of  $p$  when  $q$  and  $r$  are held constant. Here, we present the results of these simulations, together with their interpretation in the light of mathematical arguments.

Both fair offers and fair requests evolve in simulations under a wide range of mutation parameters (Figs. 2 and S2). This can be understood in the following way. Consider a resident offer  $p$  fixed in the population (plus some background deviation due to slight trembling). On one hand, if  $p$  is initially strictly lower than 0.5, responders are limiting; if a responder moves on to the second stage of pairing, she can expect to obtain a little more than  $p$  (because the most generous trembles are picked preferentially). The evolutionarily stable strategy is therefore to accept only offers at least slightly larger than  $p$  in the first stage ( $q > p$ ). This in turn generates a selective pressure toward increasing offers, because proposers must compete to attract partners. This joint increase of  $q$  and  $p$  goes on until  $p = 0.5$ , and responders are no longer limiting. On the other hand, if  $p$  is initially strictly larger than 0.5, proposers are limiting, and selection simply leads to a reduction of offers, because any offer is eventually accepted. This also goes on until  $p = 0.5$  and proposers are no longer limiting. Finally, when the resident offer is just fair ( $p = 0.5$ ), proposers and responders are in equal density, and the best strategy for responders is to accept exactly  $q = 0.5$  in the first stage, because 0.5 is exactly



**Figure 2.** Joint evolution of proposers' offer (A) and responders' request (B) in simulations, when the density of proposers and responders varies freely. Each curve is an average across 10 independent simulation runs. All parameters are like in Figure 1. In black, mutations are rare and 90% have a weak effect. In red, mutations are rare and 90% have a strong effect. In green, mutations are frequent and 90% have a weak effect. In blue, mutations are frequent and 90% have a strong effect. Trembles occur with probability 1; their effect is normally distributed with standard deviation  $5 \times 10^{-2}$ . Evolution leads to fair proposers and predominantly fair responders in all cases.

what a responder can hope to accrue in the second stage. The joint evolution of proposers and responders converges toward, and stabilizes when both sides make fair choices ( $p = q = 0.5$ ).

Note that Model 5 bears some similarity with models developed for the study of female mate choice (e.g., Janetos 1980; Real 1990); these models being themselves related to the more general study of optimal stopping rules in decision theory, for example, Chow and Robbins 1963). However, the biological foundations of the two problems significantly differ. In mate choice models, females generally do not compete for males, because each male can fertilize a very large number of females. The biological costs, and constraints, bearing upon choice are mostly a matter of time and energy consumed. In contrast, in the present model, we are interested in the consequences of competition among responders (because each proposer can only interact with one responder). The “cost of choice” we are primarily interested in is therefore what we call the “urgency” effect that emerges from competition. We show that it is sufficient to incite individuals to accept good enough offers in the first place, rather than wait indefinitely for the highest offer possible.

The energetic costs of choice should be considered in a more comprehensive version of Model 5. On one hand, as we have shown, fairness evolves when individuals are free to compare and choose their partner. On the other hand, at the opposite extreme, when switching partners is impossible like in the DG or UG, only local strategic issues matter and the most powerful partner reaps all the profits. In the intermediate case, when partner switching is possible but costly, local strategic and global market mechanisms should work together and yield an intermediate outcome. But this is outside the scope of the present article, and remains to be modeled.

## Discussion

We consider a population constituted of two trading classes, and we study the evolution of the division of a common resource when pairs of individuals, one from each class, interact. Individuals plastically choose the trading class they wish to belong to (in complementary simulations we test a different assumption where each individual's trading class is genetically encoded). Those belonging to the first class are called “proposers”; they make public offers regarding the division of the resource. Those belonging to the second class are called “responders”; they choose the most generous offer available on the market. Each individual is paired with at most one partner. We show that a “fair” division of the resource evolves. When the two classes are symmetric with respect to payoffs, evolution leads to a half-and-half division, the prime example of a fair split. When the two classes differ with respect to payoffs, the evolutionarily stable division “equalizes” the respective payoff of proposers and responders.

This result stems from a simple process based on two mechanisms: market and demography. Assume for instance that responders are initially disadvantaged by the resource partition so that their average payoff is lower than proposers'. In such a situation, playing the role of a responder is not attractive and, if one's social role is a plastic choice, responders then become scarce. This rarity in turn grants responders an advantage on the market, as they can now choose among many offers, which eventually leads to the evolution of a more favorable partition of the resource. The inverse process occurs if proposers are initially disadvantaged. Splitting the resource into two identical halves (or more generally in a way that equalizes the payoffs of the two classes) is thus the only evolutionarily stable outcome. In economic terms, the demographic feedback linking resource partition to the frequency

of each trading class allows the market to “clear” (supply and demand become equal), and this occurs when the price is just fair.

Interestingly, the same demographic feedback also takes place if social roles are genetically inherited rather than plastically chosen (an assumption undertaken in our simulations, Fig. 2A), because competition then plays the role of plastic choice. If responders, for instance, are initially disadvantaged by the resource partition, they are outcompeted by proposers and become rare, this leads in turn to the evolution of a more favorable partition of the resource, and vice versa. Our results are therefore also applicable if proposers and responders are members of two distinct species involved in a mutualism.

In the case of human behavior, the egalitarian preferences evolving under this process correspond to what is often called “inequity aversion” in behavioral economics (Fehr and Schmidt 1999). They also match the commonsense notion of fairness: Everything else being equal, individuals should be treated equally, and when everything else is not equal, distributions should compensate for any initial disequilibrium. And yet, in both the economic and the evolutionary literature, inequity aversion—or fairness—is often assumed to be a product of group selection operating at the level of cultural clusters (Boyd and Richerson 1982; Fehr and Fischbacher 2003; Choi and Bowles 2007; Haidt 2007; Bowles 2009; Boyd and Richerson 2009). Our approach offers an alternative explanation, with markedly distinct predictions. In particular, in contrast with group selection models, our results entail that human beings should be indifferent to global welfare. This is apparent, in particular, in the case of asymmetric interactions, in which the evolutionarily stable resource division predicted in our model can be far from the “social optimum” (the split that maximizes the sum of the two partners’ payoffs). At first, this result may seem surprising. And yet, as it happens, human moral intuitions are precisely surprising in this very respect (Baron 1994). For instance, people refuse to sacrifice one individual to save several others in the trolley dilemma (Greene et al. 2001), they show considerable opposition to the idea that medical treatment should be distributed according to the utility of each individual for the society at large (Marshall et al. 1999), they do not wish to increase the global wealth of a society if it is not based on fairness (Mitchell et al. 2003), etc. The predictions derived from market models are thus, at the very least, compatible with important empirical patterns, and may even help us to understand them.

The key selective lever favoring fairness in our models is partner choice, and this is empirically sound. A range of recent experiments have indeed demonstrated its importance as a mechanism to enforce cooperative behaviors in humans. People tend to select the most cooperative individuals in social encounters, and those who contribute less than others are gradually left out of cooperative exchanges (Ehrhart and Keser 1999; Sheldon

et al. 2000; Barclay 2004; Coricelli et al. 2004; Page et al. 2005; Hardy and Van Vugt 2006; Barclay and Willer 2007; Chiang 2010; Sylwester and Roberts 2010). The current proximate importance of partner choice thus strongly suggests that it might have played a major selective role in the past, leading to the evolution of fairness.

The market clearing observed in our model contrasts with what goes on in other biological markets, and in particular in biology’s oldest market paradigm: the mating market (Bowles and Hammerstein 2003; Hammerstein and Hagen 2005). Like the proposers and responders of our models, males and females interact to achieve a common good (reproduction), but the principles at work in our models apparently do not apply in mating, as evolution does not lead to the equalization of supply and demand (sperm remains produced in excess). A number of factors might explain this discrepancy (see Bowles and Hammerstein 2003). First, the division of the common good in mating is constrained by fecundation as each sex receives half the total benefits. Second, a single male can sometimes monopolize a large number of females, which proposers cannot do in our model. Third, neither reputation nor commitments are easily achieved in mating, and this limits the evolutionary stability of nuptial gifts and postzygotic paternal care. Fourth, and perhaps most important, in contrast with responders, females not only choose males in function of their offers (gift or care), but also to a large degree as a function of their sexual attractiveness: sexy males benefit, in effect, from a monopoly. All this combines to explain why mating markets do not behave in accordance with the predictions derived from our model.

In this article, we have neglected several important mechanisms that further models should investigate. For instance, our models do not offer a mechanistic account of the plastic individual decisions linking the resource division to the demography of each trading class (we simply assumed that these decisions collectively lead to the equalization of payoffs). We also did not consider the fact that the best strategy might be plastic, consisting in observing others’ offers before deciding upon one’s own, thereby generating cultural dynamics of resource partition. Most importantly, a mechanism that we neglected might be at the origin of an important discrepancy between our egalitarian predictions and psychological observations. We assumed that demographic feedback—linking the social benefits accrued by one class to the frequency of this class—operates freely. This unavoidably led to the equalization of the payoff of each class and, in economic terms, allowed the market to “clear” (i.e., offer and demand became equal). In reality, however, every social role is not accessible to everyone, at least not at the same cost, and not with the same benefits; certain roles can be more “difficult” to play. In this case, the demographic feedback does not operate freely, as certain classes remain rare for “extrinsic” reasons, and this can prevent



market clearing (see also Hammerstein and Hagen, 2005). In the same way, in an interspecific mutualism where trading classes are two species, one species might be maintained at a low (or high) density for extrinsic reasons. In either case, inherently rare trading classes could receive a larger payoff than others at evolutionary equilibrium. In the human case, this could help explain the fact that moral psychology is often at variance with egalitarianism (e.g., Marshall et al. 1999; Konow 2003).

To conclude, let us go back to the opening. We began our article by deploring the fact that most evolutionary approaches on cooperation aim to understand how cooperation can exist at all, and not what it should actually look like, and we claimed to remedy this. What we did not say, however, is that understanding how cooperation “looks” raises a difficult game-theoretic question: the question of equilibrium selection. Models of direct (or indirect) reciprocity do indeed show that cooperation is possible at equilibrium, but they also show that many different quantitative levels of cooperation are possible (in game theory, this is known as the folk theorem; see e.g., Aumann and Shapley 1994). This profound indeterminacy of equilibrium can be seen as a symptom of the fact that reciprocal cooperation (in a large sense) fundamentally raises problems of coordination, and that coordination problems often have multiple solutions. Even though equilibrium selection has primarily preoccupied theorists in economics (e.g., Harsanyi and Selten 1988), two mechanisms inspired by evolutionary theory have been proposed to select among equilibria. The first relies on the second condition of evolutionary stability (accounting for the efficiency of mutants in front of themselves: Maynard Smith and Price 1973) and tends to favor utilitarian equilibria, maximizing the sum of all payoffs (e.g., Fudenberg and Maskin 1990; Binmore and Samuelson 1992; André and Day 2007; or see also de Mazancourt and Schwartz 2010, in an explicit trading context). The second relies on group selection and also tends to favor utilitarian outcomes (Boyd and Richerson 1990). We suggest that partner choice could be a third mechanism, favoring “fair” but not necessarily utilitarian outcomes. If this intuition is valid, then it is essential to understand which equilibrium selection mechanism is most likely to be found in real-life settings, as their respective predictions differ markedly.

#### ACKNOWLEDGMENTS

We thank C. Chevallier, F. Detchevery, B. Voekl, and two anonymous reviewers for their very helpful comments on this manuscript. JBA thanks the Centre National de la Recherche Scientifique and NB the European Commission for funding.

#### LITERATURE CITED

André, J. B., and T. Day. 2007. Perfect reciprocity is the only evolutionarily stable strategy in the continuous iterated prisoner's dilemma. *J. Theor. Biol.* 247:11–22.

Aumann, R. J., and L. S. Shapley. 1994. Long-term competition: a game-

theoretic analysis. Pp. 1–15 in N. Megiddo, ed. *Essays in game theory in honor of Michael Maschler*. Springer, New York.

Axelrod, R. 1981. The emergence of cooperation among egoists. *Am. Polit. Sci. Rev.* 75:306–318.

Barclay, P. 2004. Trustworthiness and competitive altruism can also solve the “tragedy of the commons”. *Evol. Hum. Behav.* 25:209–220.

Barclay, P., and R. Willer. 2007. Partner choice creates competitive altruism in humans. *Proc. R. Soc. Lond. B* 274:749–753.

Baron, J. 1994. Nonconsequentialist Decisions. *Behav. Brain Sci.* 17:1–10.

Binmore, K., and L. Samuelson. 1992. Evolutionary stability in repeated games played by finite automata. *J. Econ. Theor.* 57:278–305.

Bowles, S. 2009. Did warfare among ancestral hunter-gatherers affect the evolution of human social behaviors? *Science* 324:1293–1298.

Bowles, S., and P. Hammerstein. 2003. Does market theory apply to biology. Pp. 153–165 in P. Hammerstein, ed. *Genetic and cultural evolution of cooperation*. MIT Press, Cambridge, MA.

Boyd, R., and P. J. Richerson. 1982. Cultural transmission and the evolution of cooperative behavior. *Hum. Ecol.* 10:325–351.

———. 1990. Group selection among alternative evolutionarily stable strategies. *J. Theor. Biol.* 145:331–342.

———. 2009. Culture and the evolution of human cooperation. *Philos. Trans. R. Soc. Lond. B* 364:3281–3288.

Camerer, C. 2003. *Behavioral game theory. Experiments in strategic interaction*. Princeton Univ. Press, Princeton, NJ.

Chiang, Y. S. 2010. Self-interested partner selection can lead to the emergence of fairness. *Evol. Hum. Behav.* 31:265–270.

Choi, J. K., and S. Bowles. 2007. The coevolution of parochial altruism and war. *Science* 318:636–640.

Chow, Y., and H. Robbins. 1963. On optimal stopping rules. *Z. Wahrscheinlichkeitstheorie* 2:33–49.

Coricelli, G., D. Fehr, and G. Fellner. 2004. Partner selection in public goods experiments. *J. Conflict Resolution* 48:356–378.

de Mazancourt, C., and M. W. Schwartz. 2010. A resource ratio theory of cooperation. *Ecol. Lett.* 13:349–359.

Ehrhart, K., and C. Keser. 1999. Mobility and Cooperation: on the Run. *Série scientifique (CIRANO):99s-24*. Available at <http://www.cirano.qc.ca/pdf/publication/99s-24.pdf>.

Fehr, E., and U. Fischbacher. 2003. The nature of human altruism. *Nature* 425:785–791.

Fehr, E., and K. Schmidt. 1999. A theory of fairness, competition, and cooperation. *Q. J. Econ.* 114:817–868.

Fudenberg, D., and E. Maskin. 1990. Evolution and cooperation in noisy repeated games. *Am. Econ. Rev.* 80:274–279.

Geritz, S. A. H., E. Kisdi, G. Meszina, and J. A. J. Metz. 1998. Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evol. Ecol.* 12:35–57.

Greene, J., R. Sommerville, L. Nystrom, J. Darley, and J. Cohen. 2001. An fMRI investigation of emotional engagement in moral judgment. *Science* 293:2105–2108.

Güth, W., R. Schmittberger, and B. Schwarze. 1982. An experimental-analysis of ultimatum bargaining. *J. Econ. Behav. Organ.* 3:367–388.

Haidt, J. 2007. The new synthesis in moral psychology. *Science* 316:998–1002.

Hammerstein, P., and E. H. Hagen. 2005. The second wave of evolutionary economics in biology. *Trends Ecol. Evol.* 20:604–609.

Hardy, C. L., and M. Van Vugt. 2006. Nice guys finish first: the competitive altruism hypothesis. *Pers. Soc. Psychol. Bull.* 32:1402–1413.

Harsanyi, J., and R. Selten. 1988. *A general theory of equilibrium selection in games*. MIT Press, Cambridge, MA and London, England.

Hauert, C., M. Holmes, and M. Doebeli. 2006. Evolutionary games and population dynamics: maintenance of cooperation in public goods

- games (vol 273, pg 2565, 2006). *Proc. R. Soc. Lond. B* 273:3131–3132.
- Janetos, A. C. 1980. Strategies of female mate choice – a theoretical-analysis. *Behav. Ecol. Sociobiol.* 7:107–112.
- Johnstone, R. A., and R. Bshary. 2008. Mutualism, market effects and partner control. *J. Evol. Biol.* 21:879–888.
- Konow, J. 2003. Which is the fairest one of all? a positive analysis of justice theories. *J. Econ. Lit.* XLI:1188–1239.
- Lehmann, L., and L. Keller. 2006. The evolution of cooperation and altruism— a general framework and a classification of models. *J. Evol. Biol.* 19:1365–1376.
- Leimar, O., and P. Hammerstein. 2001. Evolution of cooperation through indirect reciprocity. *Proc. R. Soc. Lond. B* 268:745–753.
- Marshall, G., A. Swift, D. Routh, and C. Burgoyne. 1999. What is and what ought to be: popular beliefs about distributive justice in thirteen countries. *Europ. Sociologic. Rev.* 15:349–367.
- Maynard Smith, J., and G. R. Price. 1973. The logic of animal conflict. *Nature* 246:15–18.
- Mitchell, G., P. Tetlock, D. G. Newman, and J. S. Lerner. 2003. Experiments behind the veil: structural influences on judgments of social justice. *Polit. Psychol.* 24:519–547.
- Noë, R., and P. Hammerstein. 1994. Biological markets—supply-and-demand determine the effect of partner choice in cooperation, mutualism and mating. *Behav. Ecol. Sociobiol.* 35:1–11.
- Nowak, M., and K. Sigmund. 1992. Tit for tat in heterogeneous populations. *Nature* 355:250–253.
- Nowak, M. A. 2006. Five rules for the evolution of cooperation. *Science* 314:1560–1563.
- Nowak, M. A., K. M. Page, and K. Sigmund. 2000. Fairness versus reason in the ultimatum game. *Science* 289:1773–1775.
- Nowak, M. A., and K. Sigmund. 1998. Evolution of indirect reciprocity by image scoring. *Nature* 393:573–577.
- Page, T., L. Putterman, and B. Unel. 2005. Voluntary association in public goods experiments: reciprocity, mimicry and efficiency. *Economic J.* 115:1032–1053.
- Real, L. 1990. Search theory and mate choice. I. models of single-sex discrimination. *Am. Nat.* 136:376–405.
- Rubinstein, A. 1982. Perfect equilibrium in a bargaining model. *Econometrica* 50:97–110.
- Schelling, T. C. 1960. *The strategy of conflict*. Harvard Univ. Press, Cambridge.
- Sheldon, K. M., M. S. Sheldon, and R. Osbaldiston. 2000. Prosocial values and group assortment—within an N-person prisoner’s dilemma game. *Hum. Nature Int. Bios.* 11:387–404.
- Stahl, I. 1972. *Bargaining theory*. Economics Research Institute, Stockholm.
- Sylwester, K., and G. Roberts. 2010. Cooperators benefit through reputation-based partner choice in economic games. *Biol. Lett.* 6:659–662.

Associate Editor: M. Doebeli

## Supporting Information

The following supporting information is available for this article:

**Figure S1.** Evolution of the average offer in simulations, when the density of proposers and responders vary freely.

**Figure S2.** Joint evolution of proposers’ offer (A) and responders’ request (B), when the density of proposers and responders vary freely.

Supporting Information may be found in the online version of this article.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.