# Evolution of equal division among unequal partners

**Stéphane Debove**[1,2*], **Nicolas Baumard**[3], **Jean-Baptiste André**[1]

[1]Institut de Biologie de l'Ecole normale supérieure (IBENS), INSERM 1024, CNRS 8197, Ecole normale supérieure - PSL Research University, Paris, France.

[2]Université Paris Descartes, Paris, France.

[3]Institut Jean-Nicod (CNRS - EHESS - ENS), Département d'Etudes Cognitives, Ecole normale supérieure - PSL Research University, Paris, France.

*__Correspondence__: Stéphane DEBOVE, Institut de Biologie de l'Ecole normale supérieure, 46 rue d'Ulm, 75005 Paris, FRANCE. Email address: sd@stephanedebove.net

Data available from the Dryad Digital Repository: http://doi.org/10.5061/dryad.36pn8 and on the first author's website: http://stephanedebove.net/?p=44

The definitive version was published in Evolution, 69: 561–569. doi: 10.1111/evo.12583

## Abstract

One of the hallmarks of human fairness is its insensitivity to power: while strong individuals are often in a position to coerce weak individuals, fairness requires them to share the benefits of cooperation equally. The existence of such egalitarianism is poorly explained by current evolutionary models. We present a model based on cooperation and partner choice that can account for the emergence of a psychological disposition toward fairness, whatever the balance of power between the cooperative partners. We model the evolution of the division of a benefit in an interaction similar to an ultimatum game, in a population made up of individuals of variable strength. The model shows that strong individuals will not receive any advantage from their strength, instead having to share the benefits of cooperation equally with weak individuals at the evolutionary equilibrium, a result that is robust to variations in population size and the proportion of weak individuals. We discuss how this model suggests an explanation for why egalitarian behaviors towards everyone, including the weak, should be more likely to evolve in humans than in any other species.

**Keywords:** Egalitarianism, Fairness, Partner Choice, Game Theory.

# 1 Introduction

For centuries, philosophers have discussed and denounced the fallacy of the "law of the strongest." In the first chapter of Plato's *Republic*, for instance, Thrasymachus claims that "justice is nothing else than the interest of the stronger," which Socrates then disputes. Many years later, in his foundational work on political rights, Rousseau noted that he could not see "how morality could result from the effects of physical power" (Rousseau, 1762). In the last decades, scientific research has accumulated evidence of humans' strong preference for egalitarian outcomes independently of the power relationship between individuals (Fehr and Schmidt, 1999; Fehr and Fischbacher, 2003; Camerer, 2003; Boehm, 1993, 1997; Dawes et al., 2007; Tricomi et al., 2010). In economic games, it has been shown that in collaboratively meaningful contexts, people favor equal divisions when contributions are equal (Dawes et al., 2007; Cappelen et al., 2007; Frohlich et al., 2004). A similar phenomenon is detectable early on in children, with three-year-olds dividing rewards equally after collaboration (Warneken et al., 2011). Finally, cross-cultural studies have shown that such behavior can be observed in many different cultural contexts, from small-scale societies to large industrial societies. In fact, this phenomenon is so universal that anthropologists have coined the term "egalitarian syndrome" to describe the prevalence of such preferences for equality in small-scale societies (Boehm, 1993, 1997; Cashdan, 1980). More generally, modern mass movements such as the anti-slavery, anti-discrimination, civil rights, and fair trade movements are all expressions of the same urge to care for and defend the interests of the weak.

Evolutionarily speaking, these observations raise the question: why should natural selection favor equal divisions of benefits, independently of the power struggle between the protagonists? Or, said differently, under what conditions is it adaptive for stronger or dominant individuals to leave half of the resource to their partners, when they could keep everything for themselves?

A useful paradigm for studying this question is the ultimatum game (UG). In this game, two individuals bargain over the division of a benefit, with one individual (the "proposer") making an offer to the other individual (the "responder"). If the responder accepts the offer, it is implemented; otherwise, none of the individuals receives any benefit. The very structure of this game implies an asymmetry of bargaining power between the two players. On the one hand, whatever offer a responder's partner makes, accepting it brings a greater gain than rejecting it. Therefore, in all cases, natural selection favors indiscriminate responding, with responders taking whatever benefits are made available to them. On the other hand, and as a result, selection favors stingy proposers, offering the minimal possible amount. The division of benefits at the evolutionary equilibrium is thus maximally "unfair": the empowered individual (the proposer) keeps virtually all the benefits. The UG is therefore a conservative paradigm for studying the evolution of fairness.

In asymmetric interactions of this sort, in which a dominant individual can unilaterally impose a division of resources on another, both bargaining theory (in economics) and reproductive skew theory (in behavioral ecology) show that the dominated individual's outside options limit the level of inequality that the dominant can impose (Vehrencamp, 1983; Johnstone, 2000; Muthoo, 1999). However, these models treat the value of outside options as *exogenous* parameters, which are fixed a priori. Hence, whereas they can account for the fact that dominant individuals leave "something" to subordinates, they cannot explain quantitatively why *equal* divisions should precisely be favored by natural selection. André and Baumard (2011a) went a step further by showing that if outside options consist in the possibility of entering into another identical interaction with a new partner, and if the dominant or dominated status of an individual is randomly decided in each interaction, then each individual is certain to receive an expected payoff equal to half of the resource in the next interaction she will enter (see also André and Baumard (2011b) for a different model leading to the same consequence). This outside option thus forces dominant individuals to always share benefits in two equal parts, a mechanism that was already suggested verbally by Vehrencamp (1983) and that is also conceptually similar to the infinite-horizon, alternating-offers bargaining game of Rubinstein (1982).

André and Baumard (2011a)'s result relies heavily on the assumption that the dominant or subordinate status of an individual is randomly decided in each new interaction. In other words, in this scenario, the equal division of resources is the result of equal outside options. But the biological plausibility of equal outside options is highly debatable. For instance, in any scenario in which dominance is not random but linked to some intrinsic property of the individual, such as physical strength, individuals who are dominant in an interaction will also be more likely to be dominant in other interactions. This means that two individuals engaged in an interaction will not have the same outside options. Hence, it is legitimate to wonder whether the evolution of equal divisions will hold after the introduction of asymmetries of strength among individuals. If not, the observation that humans share equally even with weaker individuals, a central characteristic of fairness, would require another explanation.

Note that other scholars have proposed alternative explanations for the evolution of human fairness not based on the possibility of changing partners (Nowak et al., 2000; Gale et al., 1995; Rand et al., 2013). We will present these explanations in the discussion and compare them with our own approach, but at this stage it is important to note that none of these alternative approaches takes into account possible asymmetries of strength between bargaining individuals. To our knowledge, this paper is the first theoretical study concerned with the evolution of human fairness that explicitly considers systematic asymmetries of strength. In the discussion, we will also highlight the limits of our model and its relationship with the non-human literature on biological markets (e.g., Noë et al., 1991; Noë and Hammerstein, 1995; Noë et al., 2001).

Here, we present both an analytical model and the results of individual-based simulations on the evolution of the division of a benefit in an ultimatum game-like interaction, in a population where individuals can change social partner. Individuals are assumed to be characterized by an intrinsic "strength" that affects their probability of playing the strategically dominant role of proposer in all UGs that they play. We investigated whether fair divisions evolve in such an environment, and in particular whether strong individuals refrain from taking advantage of their strength when they are paired with a weaker partner.

# 2 Methods

**Individual-based simulations.**

We consider a population of individuals who enter into a series of pairwise social interactions with random partners. All individuals begin their lives in a solitary state, and they then meet random social partners, among other solitary individuals, at a given constant rate $\beta$. When two individuals meet, one of them is given the role of proposer, with a probability that may depend on the relative strength of the two individuals (see below). The proposer offers a given division of benefits to the responder, who can then either accept or refuse. If the offer is accepted, the two individuals actually enter into the social interaction, which is assumed to take time. Hence, they leave the pool of available solitary individuals until the end of their interaction, which occurs at a constant "split" rate $\tau$. On the other hand, if the offer is rejected, the two individuals immediately go back to the pool of solitary individuals without receiving any benefit. Note that although it is convenient to describe this interaction as an UG, it is not a real UG stricto sensu, as the responder always has the choice of refusing an offer and hoping to interact with someone else in the population, which is not the case in the UG.

We consider a Wright-Fisher population with non-overlapping generations. Each generation lasts for a constant number of time steps, at which point all individuals reproduce according to the amount of benefits they have accumulated throughout their life, and then die. Genetic recombination is allowed between each generation.

*The cost of partner choice.*

The cost of partner choice is implicit in the above model. It is a consequence of the time it takes to find a new partner after the rejection of an offer. Hence, the cost and benefit of being choosy are not controlled by explicit parameters, but by two parameters that characterize the "fluidity" of the social market: the "encounter rate" $\beta$, and the "split rate" $\tau$. When $\frac{\beta}{\tau}$ is large, interactions last a long time (low split rate $\tau$) but finding a novel partner is fast (high encounter rate $\beta$), and individuals thus should be picky about which offers they accept. On the contrary, when $\frac{\beta}{\tau}$ is low, interactions are brief but finding a novel partner takes time, and individuals should thus accept almost any offer.

*Strength.*

We assume individuals are characterized by an intrinsic quantitative property $\sigma \in [0,1]$ representing their "strength", which affects their probability of playing the advantageous role of proposer in the UG. This intrinsic property is constant across the entire life of an individual but is not heritable: i.e., at birth, each individual is randomly attributed a given level of "strength" that is independent of his parent's, according to a random distribution (see below). In an interaction between two individuals, we assume that the strongest of the pair has a given constant probability $\frac{1}{2} * (1+\phi)$ of playing the role of proposer, where $\phi \in [0,1]$ is a constant parameter, independent of the quantitative difference between the partners' strengths. When $\phi = 1$, strength controls the attribution of roles deterministically: the stronger partner always plays the role of proposer. When $\phi = 0$, strength has no effect on the assignment of roles. We also assume that, when two individuals of *exactly* equal strength are paired together, they have an equal chance of playing the role of proposer.

Regarding the distribution of individual strength at birth, for the sake of simplicity in our analytical approach and in most of our simulations, we assume that there are only two strengths, and thus only two types of individuals ("strong" and "weak"). In this case, we will call any given individual's probability of being randomly designated as "weak" at birth $x$. In other versions of our simulations, we assume instead that the strength of an individual at birth is sampled from a uniform distribution between 0 and 1. In this case, individuals never interact with a partner of the exact same strength.

*The social strategy*

To play a UG, each individual must be characterized by two different behavioral variables: the *offer* they make when they play the role of proposer, and their *request* as a responder, i.e., the minimum offer they are ready to accept from their partner. The aim here is to consider the possibility of individuals detecting their partner's strength and adapting their behavior accordingly.

With only two levels of strength in the population, we assume that individuals are characterized by eight genetic variables: four $p_{ij}$ and four $q_{ij}$ variables, with $i$ and $j \in \{s,w\}$ denoting an individual's strength ($s$ for "strong", $w$ for "weak"). $p_{ij}$ is the offer made by a proposer of strength $i$ in an interaction with a responder of strength $j$. $q_{ij}$ is the minimum offer that a responder of strength $i$ is ready to accept in an interaction with a proposer of strength $j$. For example, a strong individual who is the proposer in an interaction with a weak individual will propose $p_{sw}$ benefits to the responder. The weak individual will then compare the value of $p_{sw}$ to his own $q_{ws}$, and if $p_{sw} \geq q_{ws}$, the offer will be accepted.

With a continuum of strength in the population, we assume the offer is controlled by three underlying genetic traits: a constant $\gamma$, a degree of linear dependence on the individual's own strength $\rho_p$, and a degree of linear dependence on the partner's strength $\rho_r$. The offer is given by:

$$Offer = \gamma + \rho_p * \sigma_p + \rho_r * \sigma_r$$

with $\sigma_p$ being the strength of the proposer, $\sigma_r$ being the strength of the responder, $\gamma \in [0,1]$ and $\rho_p, \rho_r \in [0,1]$. Correspondingly, with a continuum of strength the responder's request is genetically encoded by three loci $\mu$, $\lambda_p$ and $\lambda_r$, and given by the expression:

$$Request = \mu + \lambda_p * \sigma_p + \lambda_r * \sigma_r$$

with $\sigma_p$ being the strength of the proposer, $\sigma_r$ being the strength of the responder, $\mu \in [0,1]$ and $\lambda_p, \lambda_r \in [0,1]$.

Note that although the system of offers and requests is a convenient way to model these interactions, it can be interpreted biologically in a different and probably more realistic way. The existence of offers does not necessarily imply an underlying contract: offers can also mean that responders have some information on the proposer's usual behaviors, for example through a reputation built up in the course of past interactions with other individuals. Therefore, when an individual is characterized by an offer *p*, we can also interpret this as this individual having the public "reputation" of offering *p*.

**Analytical model.**

The analytical model incorporates all of the features of the simulations presented above, but with one simplification: we assume that the total number of interactions accepted per unit of time is the same for each individual. With this assumption, rejecting an opportunity to cooperate does not compromise the chances of cooperating later, but on the contrary grants new opportunities. This situation is analogous to the condition where $\frac{\beta}{\tau}$ tends towards infinity in the simulations: social opportunities are plentiful at the scale of the length of interactions. When individuals reject an interaction, however, they are forced to postpone their social interaction to a later encounter. We assume that this entails an explicit cost expressed as a discounting factor $\delta$ ($0 \leq \delta < 1$). If we call the average payoff of an individual of strength $i$ $G_i$, then $\delta G_i$ will be the average expected payoff in the next interaction after rejecting an offer. When $\delta$ equals 1, refusing an interaction carries no cost; when $\delta$ equals 0, refusing an offer will result in zero payoff from the next interaction. In practice, we will neglect the case where $\delta$ equals 1, as it leads to artefactual results (see SM section 1.2). The analytical model is fully explained and solved in SM section 1.2.

The question we want to answer is the following: how will offers and requests evolve in such an environment, where individuals of different strengths coexist and share resources? Unless otherwise specified, the following results are concerned with the case where there are only two strengths coexisting in the population.

# 3 Results

When the population is made up of equal numbers of strong and weak individuals ($x = \frac{1}{2}$), if partner choice is not costly, the difference in strength between strong and weak individuals has little impact on the offers that strong individuals make to weak individuals (Fig. 1). If partner choice is not costly, starting from a stingy population of strong individuals offering nothing to weak individuals, offers progressively raise in the population up to the point where the strong offer close to half of the resource to their weak partners (Fig. 1 circle markers). In fact, strong individuals offer weak individuals as much benefit as they offer to other strong individuals when partner choice is not costly (SM Fig.1). On the contrary, when partner choice is highly costly (Fig. 1 diamond markers), strong individuals make very low offers to weak individuals. This result holds even if there is a continuum of strengths in the population (not just two; SM Fig. 2). As long as partner choice is not too costly, at the evolutionary equilibrium individuals who are paired with a stronger individual and playing the role of responder will still receive half of the total resource to be shared.

In the previous results, we arbitrarily set the proportion of weak individuals ($x$) at 0.5. It is plausible to think that this parameter will influence the division of benefits, since it impacts the social opportunities of weak individuals. To determine if equal divisions can still evolve when there is a low proportion of weak individuals in the population, we ran simulations for different values of $x$. The results show that this parameter in fact has a very limited impact: divisions of resources between strong and weak individuals continue to be equal when the percentage of weak individuals is as low as 5% (Fig. 2 left panel).

Population size is another parameter that could affect the payoff to the weak: the smaller the population size, the smaller the total amount of social opportunities available to each individual. To test the effect of this parameter, we analyzed the payoff to weak individuals at the evolutionary equilibrium in populations of different sizes. Population size plays a role in determining the payoffs received by weak individuals, but quasi-equal divisions can be found in populations as small as 50 individuals (Fig. 2 right panel).

The results of the analytical model confirm the simulation results. When partner choice is not costly, at the evolutionary equilibrium strong individuals do not take advantage of their strength to offer unequal divisions to weak individuals (Fig. 3). Even in the case where weak individuals are *always* in the strategically dominated position of responder when paired with a strong individual ($\phi = 1$), they will receive half of the resource at the equilibrium as long as the cost of changing partners is not too high.

Analytical results also confirm that the frequency of weak individuals in the population has a small impact on divisions at the evolutionary equilibrium. Figure 4 shows the offer made by a strong individual to a weak individual at the evolutionary equilibrium, when $\phi = 1$, for different values of $x$. As long as partner choice is not too costly, weak individuals will receive close to half of the resource to be shared, even when there are not many of them in the population ($x \to 0$). However, the higher the cost of changing partners, the more restrictive the parameter $x$ becomes.

## 4 Discussion

In this article, we have shown that equal divisions can evolve in an interaction similar to the ultimatum game even when some individuals are stronger than others in the population, and thus have better average outside options than other individuals. Although they have a strategic bargaining advantage, strong individuals agree to give close to half of the benefits of interactions to weak individuals at the evolutionary equilibrium, a result that is robust to variations in population size and in the proportion of strong individuals in the population. To our knowledge, this is the first theoretical study on the evolution of human fairness that explicitly considers systematic asymmetries of strength, and shows that strength is not an evolutionarily relevant parameter to determine the division of benefits in an environment where partner choice is possible. In particular, we relax a crucial assumption used in previous modelling approaches (André and Baumard, 2011a) to show that equal divisions can still evolve when some individuals have better average social opportunities than others, a condition necessary for understanding the reach of partner choice-based fairness in humans.

These results shed light on an interesting question: why do strong individuals not more often take advantage of their strength to exploit weak individuals? The answer seems to be that the advantage of strength is only *local*, when what matters when individuals can choose their social partners are *global* social opportunities. When individuals are embedded in a rich network of cooperative interactions and social opportunities, their bargaining power is determined not by local dominance relationships in each interaction, but by their outside options in the population as a whole. Thus, a weak individual's bargaining power is not affected by being locally dominated in a pairing with a strong individual, because of all the social opportunities that are available with other partners. Confronted with an unfair offer from a strong individual, a weak individual can easily refuse it and wait for an encounter with another weak individual. As a consequence, a strong individual who wants to interact with a weak individual will have to offer at least what the weak individual could gain elsewhere.

It is important to note that this result holds even when the relative proportion of weak individuals in the population is small. Variation in the number of weak individuals does indeed affect the average social opportunities of the weak. However, as long as a weak individual has at least one other weak partner in the population, this potential cooperative opportunity will constrain what strong individuals can offer. This result highlights the fact that the factor that most determines individuals' payoffs in an environment with varied social opportunities is not the *average* of the social opportunities that are available to them, but their *best* social opportunities. Weak individuals can "put forward" their best social opportunities, in which they can gain $\frac{1}{2}$ on average, when they are bargaining with strong individuals, which leads to the evolution of equal divisions even in strongly unbalanced populations.

Note that our results could also be interpreted the other way around: if partner choice is not sustained, the evolutionarily stable strategy will not be fair. If human beings have the ability to adapt (plastically) their level of fairness to cues indicating the efficiency of partner choice, then this could help to explain some of the inequalities we observe in pastoralist or agriculturalist societies, going from mild stratification to extreme cases of despotism and slavery (Kaplan et al., 2009; Summers, 2005), or inequalities in modern societies, for example in situations of monopoly (Kahneman et al., 1986; Piketty and Saez, 2014). Although these links are purely speculative at this point, someone interested in the history of inequalities could make interesting predictions with our model.

Our model is related to existing models of the evolution of human fairness, but previous models have not taken into account differences of strength between individuals, and thus cannot explain the emergence of equal divisions when power is unequally distributed. A widely-cited model by Nowak et al. (2000) suffers from a restriction of the parameter space that undermines its main result: offers progressively increase because responders can build up a reputation for refusing low offers, but offers stabilize at 50% only because the authors arbitrarily assume that proposers cannot request more as a responder than what they offer as a proposer, following an unsubstantiated assertion that "individuals do not regard the role of proposer as inferior to the role of responder" (Nowak et al. (2000) footnote 14; see André and Baumard (2011b) for a lengthy discussion of this problem). Another line of models shows that introducing high variance in the responders' acceptance threshold can also lead to increased offers in the ultimatum game (Gale et al., 1995; Rand et al., 2013; Ichinose, 2012). Variance, depending on the model, can be due to a high mutation rate, weak selection, learning mistakes, difficulty assessing the strategies of others, etc., but the general mechanism is the same: when noise leads a sufficiently large proportion of responders to keep refusing low offers, proposers have no choice but to increase their offers. A recent work also suggests that "spiteful" strategies negatively assorted with a mix of other strategies could lead to a certain degree of fairness (Forber and Smead, 2014), even though the use of the word "spite" in an evolutionary context is highly debatable (West and Gardner, 2010). To our knowledge, the only existing model on the evolution of equality with differences of strength is the model developed by Gavrilets (2012). In this model, the need for individuals to not only maximize their own fecundity but also minimize the fecundity of others can lead to the evolution of helping behaviors directed towards weak individuals engaged in agonistic ("owner-bully") interactions with stronger individuals. This last mechanism based on intense inter-individual competition is an alternative to the one we suggest based on intense cooperation, and it would be interesting to see how they compare when it comes to explaining finer-grained properties of human fairness (Konow, 2000; Baumard et al., 2013).

Overall, the scarcity of models of the evolution of human fairness incorporating differences of strength makes our approach more closely related to models of reproductive skew. Skew theory aims to explain why in some species the benefits of reproduction are highly skewed in favor of dominant individuals, whereas in other species a more equal division occurs and subordinates reproduce as much as dominants (Johnstone, 2000). Some "tug-of-war" models of reproductive skew have been applied to human cooperation (Barker et al., 2012), but only transactional models emphasize the important role of subordinates' outside options in determining how reproduction is divided. They show that when subordinates have good outside options, dominants are forced to give them a large number of reproductive opportunities if they want to keep them in their group (Vehrencamp, 1983; Reeve et al., 1998). Theoretically, we depart from reproductive skew models in the way we model outside options: rather than arbitrarily fixing a certain value for them, outside options in our model emerge from the dynamics of social interactions themselves, i.e., from encounters between individuals and the cost of changing partners. In this situation, if partner choice is not costly, the model shows that strong individuals not only give *something* to weak individuals, they give exactly half of the resource to be shared. Our model is also strongly inspired from biological markets models (Noë et al., 1991; Noë and Hammerstein, 1994), in which commodities are being exchanged and trading partners compete to be chosen by the other trading class. A general result of this line of models, reminiscent of our own results, is that supply and demand represented by the trading classes will determine the value of the exchanged commodity (Noë and Hammerstein, 1995; Noë et al., 2001; Johnstone and Bshary, 2008). An important difference with our model though is that we do not suppose an agent has to be firmly attached to a specific trading class. Whereas it makes sense when modeling mating markets or interspecific mutualisms to assign a fixed class to each agent (male or female, species A or species B, breeder or helper, etc.), here we investigated what happens when individuals can freely switch from one class to the other.

One important question concerns the biological interpretation of the "strength" postulated in our models. Because we chose to model strength abstractly, it can represent any feature that affects individuals' "resource-holding potential" as described classically in behavioral ecology (Arnott and Elwood, 2009): body size, body mass, development of weaponry, physiological state, etc. The central property of strength in our models, however, is that it brings individuals only a local bargaining power advantage, and not a global one. Fairness evolves in spite of asymmetries of strength if stronger individuals are more likely to dominate each interaction, but not if they can actively reduce the outside options of weaker individuals by preventing them from cooperating with each other. This definition of strength constitutes a clear limit to our model: if we assumed that stronger individuals had the possibility to coerce their partner into interacting, fairness could not evolve. However, reducing others' outside options requires a much higher investment than just taking away resources, as individuals need to monitor who else their partners are interacting with at any given moment. To do so, strong individuals would have to spend the majority of their time guarding others, to the detriment of the production of cooperative benefits. At the very least, we predict that, in situations in which such partner guarding is impossible, the benefits of cooperation should be divided equally, irrespective of the resource-holding potential of individuals. Another limit of our model is that we suppose strength is not heritable (i.e. we assume that investment into competitive strength cannot evolve by natural selection). There is no doubt it would be fruitful to relax this simplifying assumption in further models to see whether an interesting pattern of coevolution between strength and fairness emerges.

While cooperation and partner choice are not restricted to humans, egalitarianism seems to be rare outside the human species. Chimpanzees, unlike children for instance, do not share benefits equally even when they had to collaborate to produce them (Melis et al., 2006), and rarely share food at all in natural settings (Tomasello et al., 2012). Both within and beyond the primate order, high-ranking males usually enjoy more resources than low-ranking males (Ellis, 1995). In the kingdom Animalia, contests over resources are most often won by the individual with the highest resource-holding power (Arnott and Elwood, 2009). This raises the question: why are humans so prone to respect the interests of the weak? Although our model does not allow us to answer this question with certainty, it offers at least two different hints. The first is that weak individuals in the human species may have a better choice of cooperative partners than those in other species. The second hint is that in humans, strength may have a far lesser role in the generation of benefits than it is in other species. Because of the nature of human cooperation and the variety of forms it can take, there are ways for two physically weaker individuals to produce benefits equivalent to one weak and one strong individual working together (Wiessner, 1996; Kaplan et al., 2009). In other words, strength ceases to be an important factor in determining the *division* of benefits because it ceases to be an important factor in determining the *production* of benefits. In other species, on the contrary, the involvement of a strong individual almost always brings extra benefits that cannot be produced in interactions involving only weak individuals (protection, access to females, territory, or food). As a consequence, distributions skewed in favour of strong individuals are much more frequently observed in non-human animals (Ellis, 1995; Grafen, 1987). Although at this point these two non-exclusive hypotheses are speculative and cannot be confirmed theoretically, our model suggests that both the quantity of human cooperation (meaning that individuals always have a rich overall set of social opportunities) and its nature (allowing even physically weaker and hence less competitive individuals to produce similar benefits) are human-specific selection pressures that could have led to the evolution of concern for the interests of the weak. Of course, human egalitarian behaviors have varied greatly across time and space, and our simple model based on genetic transmission alone cannot capture the full complexity of human behavior and cognition in relation to cooperation. It may nonetheless be able to explain the general and universal pattern of egalitarian behaviors observed in many human societies (Boehm, 1993, 1997; Cashdan, 1980).

# References

André, J.-B. and N. Baumard, 2011a. Social opportunities and the evolution of fairness. Journal of theoretical biology 289:128–35.

André, J.-B. and N. Baumard, 2011b. The evolution of fairness in a biological market. Evolution 65:1447–56.

Arnott, G. and R. W. Elwood, 2009. Assessment of fighting ability in animal contests. Animal Behaviour 77:991–1004.

Barker, J. L., P. Barclay, and H. K. Reeve, 2012. Within-group competition reduces cooperation and payoffs in human groups. Behavioral Ecology 23:735–741.

Baumard, N., J. André, and D. Sperber, 2013. A mutualistic approach to morality: The evolution of fairness by partner choice. Behavioral and Brain Sciences 6:59–122.

Boehm, C., 1993. Egalitarian behavior and reverse dominance hierarchy. Current Anthropology 34:227–254.

Boehm, C., 1997. Impact of the human egalitarian syndrome on Darwinian selection mechanics. American Naturalist 150:S100–S121.

Camerer, C., 2003. Behavioral game theory: Experiments in strategic interaction, vol. 32. Princeton University Press.

Cappelen, A., A. Hole, E. Sø rensen, and B. Tungodden, 2007. The pluralism of fairness ideals: An experimental approach. American Economic Review 97:818–827.

Cashdan, E., 1980. Egalitarianism among hunters and gatherers. American Anthropologist 82:116–120.

Dawes, C. T., J. H. Fowler, T. Johnson, R. Mcelreath, and O. Smirnov, 2007. Egalitarian motives in humans. Nature 446:794–796.

Ellis, L., 1995. Dominance and Reproductive Success Among Nonhuman Animals: A Cross-Species Comparison. Ethology and sociobiology 333:257–333.

Fehr, E. and U. Fischbacher, 2003. The nature of human altruism. Nature 425:785–91.

Fehr, E. and K. Schmidt, 1999. A theory of fairness, competition, and cooperation. The quarterly journal of economics Pp. 817–868.

Forber, P. and R. Smead, 2014. The evolution of fairness through spite. Proceedings of the Royal Society B 281.

Frohlich, N., J. Oppenheimer, and A. Kurki, 2004. Modeling other-regarding preferences and an experimental test. Public Choice 119:91–117.

Gale, J., K. Binmore, and L. Samuelson, 1995. Learning to be imperfect: The ultimatum game. Games and Economic Behavior 8:56–90.

Gavrilets, S., 2012. On the evolutionary origins of the egalitarian syndrome. Proceedings of the National Academy of Sciences of the United States of America 109:14069–74.

Grafen, A., 1987. The logic of divisively asymmetric contests: respect for ownership and the desperado effect. Animal Behaviour 35:462–467.

Ichinose, G., 2012. Coevolution of Role Preference and Fairness. Complexity 00:1–9.

Johnstone, R. and R. Bshary, 2008. Mutualism, market effects and partner control. Journal of evolutionary biology 21:879–888.

Johnstone, R. A., 2000. Models of reproductive skew: A review and synthesis. Ethology 106:5–26.

Kahneman, D., J. Knetsch, and R. Thaler, 1986. Fairness as a constraint on profit seeking: Entitlements in the market. The American economic review 76:728–741.

Kaplan, H. S., P. L. Hooper, and M. Gurven, 2009. The evolutionary and ecological roots of human social organization. Philosophical transactions of the Royal Society B. 364:3289–99.

Konow, J., 2000. Fair shares: Accountability and cognitive dissonance in allocation decisions. The American Economic Review 90:1072–92.

Melis, A. A. P., B. Hare, and M. Tomasello, 2006. Engineering cooperation in chimpanzees: tolerance constraints on cooperation. Animal Behaviour 72:275–286.

Muthoo, A., 1999. Bargaining Theory with Applications. Cambridge University Press.

Noë, R. and P. Hammerstein, 1994. Biological markets: supply and demand determine the effect of partner choice in cooperation, mutualism and mating. Behavioral ecology and sociobiology Pp. 1–11.

Noë, R. and P. Hammerstein, 1995. Biological markets. Trends in ecology & evolution 10:336–9.

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

Noë, R., C. Schaik, and J. Hooff, 1991. The market effect: An explanation for pay-off asymmetries among collaborating animals. Ethology 87:97–118.

Nowak, M. a., K. M. Page, and K. Sigmund, 2000. Fairness versus reason in the ultimatum game. Science 289:1773–5.

Piketty, T. and E. Saez, 2014. Inequality in the long run. Science 344.

Rand, D. G., C. E. Tarnita, H. Ohtsuki, and M. a. Nowak, 2013. Evolution of fairness in the one-shot anonymous Ultimatum Game. Proceedings of the National Academy of Sciences of the United States of America 110:2581–6.

Reeve, H., S. Emlen, and L. Keller, 1998. Reproductive sharing in animal societies: reproductive incentives or incomplete control by dominant breeders? Behavioral Ecology 9:267–278.

Rousseau, J.-J., 1762. Of The Social Contract, Or Principles of Political Right.

Rubinstein, A., 1982. Perfect equilibrium in a bargaining model. Econometrica: Journal of the Econometric Society .

Summers, K., 2005. The evolutionary ecology of despotism. Evolution and Human Behavior 26:106–135.

Tomasello, M., A. P. A. 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.

Tricomi, E., A. Rangel, C. Camerer, and J. O'Doherty, 2010. Neural evidence for inequality-averse social preferences. Nature 463:1089–1091.

Vehrencamp, S. L., 1983. Optimal Degree of Skew in Cooperative Societies. American Zoologist 23:327–335.

Warneken, F., K. Lohse, A. P. Melis, and M. Tomasello, 2011. Young children share the spoils after collaboration. Psychological science 22:267–73.

West, S. a. and A. Gardner, 2010. Altruism, spite, and greenbeards. Science (New York, N.Y.) 327:1341–4.

Wiessner, P., 1996. Leveling the hunter: constraints on the status quest in foraging societies. Pp. 171–192, *in* Food and the Status Quest: An Interdisciplinary Perspective, p. wiessne ed.

# 5 Acknowledgments

SD thanks the *Région Ile-de-France* for funding this research through a *2012 DIM "Problématiques transversales aux systèmes complexes"* grant, and the *Institut des systèmes complexes* and the *Frontières du Vivant ED 474* PhD Program for their support. This work was supported by ANR-10-LABX-0087 IEC and ANR-10-IDEX-0001-02 PSL*. We thank three anonymous reviewers for their helpful comments. The data and code for the simulations are available from the Dryad Digital Repository: http://doi.org/10.5061/dryad.36pn8. They are also available on request from the first author and on the first author's website: http://stephanedebove.net/?p=44 . The authors declare no conflict of interest.

# 6 Figures

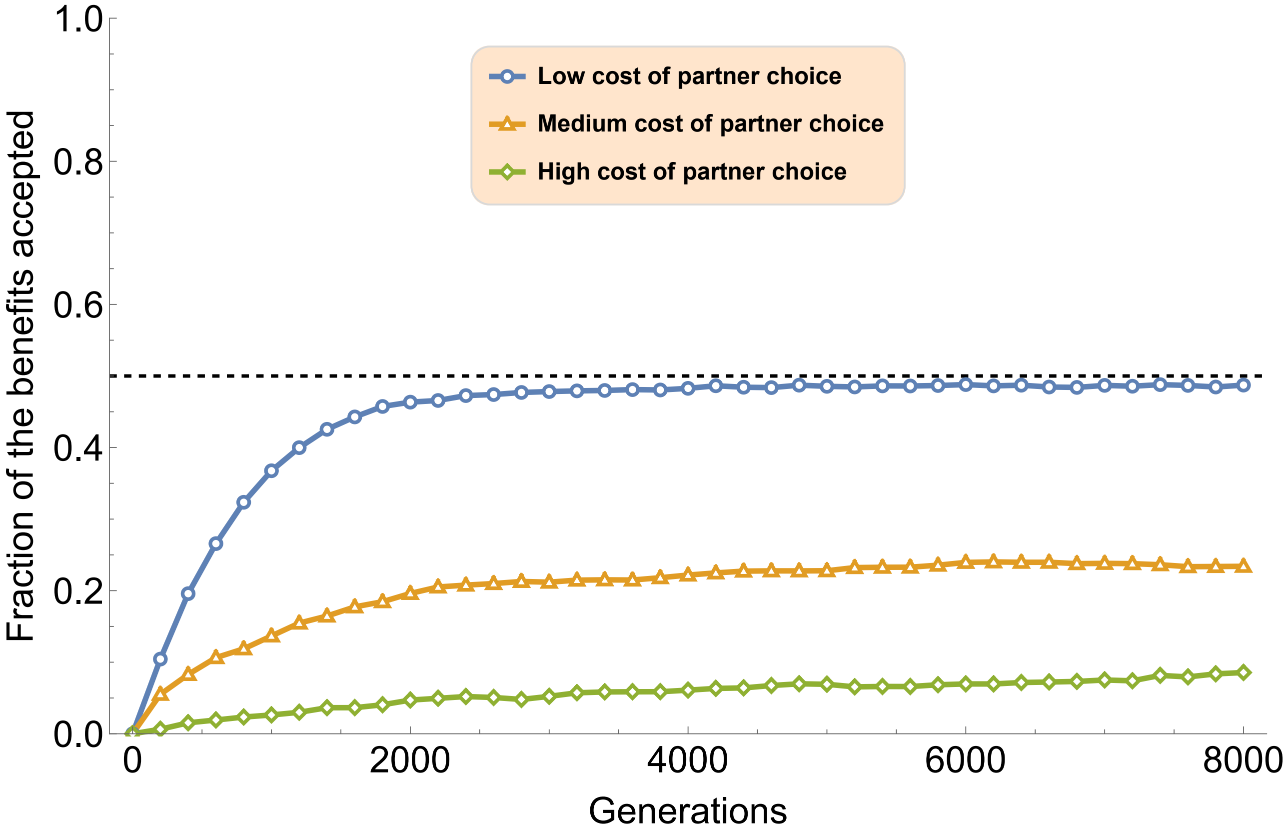

Figure 1: Average offer accepted by a weak individual paired with a strong individual across generations (simulation results). Average over 30 simulations. The dashed line corresponds to the (theoretical) perfectly equal division (50 %). Parameter values: $\phi = 1, x = \frac{1}{2}$. Circle markers: partner choice has almost no cost ($\frac{\beta}{\tau} = 100$). Triangle markers: partner choice has a medium cost ($\frac{\beta}{\tau} = 1$). Diamond markers: partner choice is highly costly ($\frac{\beta}{\tau} = 0.01$). When partner choice is not costly, weak individuals only accept offers that are close to 50% at the equilibrium. Other parameters used for these simulations can be found in SM section 1.6.

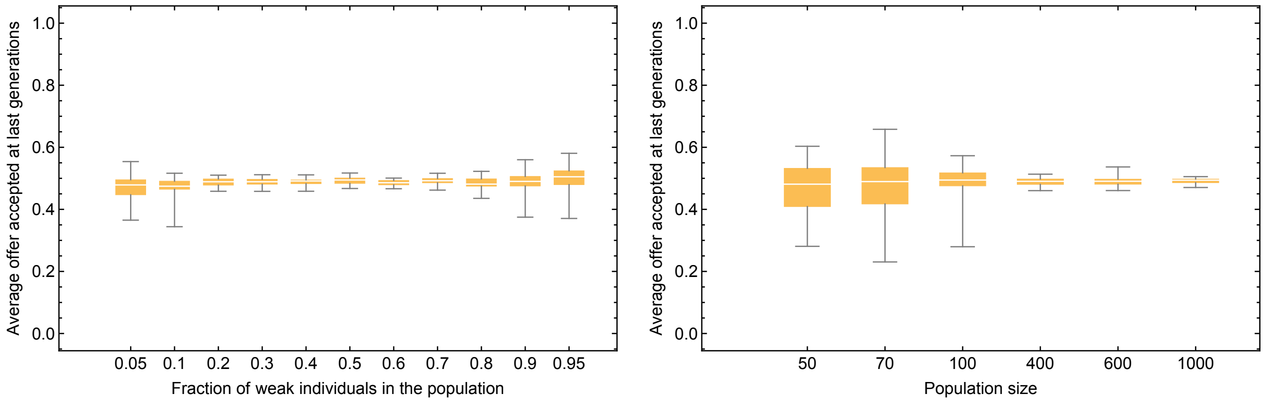

Figure 2: Robustness of the evolution of equal offers between strong and weak individuals (simulation results). Left panel: average offer accepted by a weak individual paired with a strong individual at the evolutionary equilibrium, for different fractions of weak individuals in the population. Right panel: average offer accepted by a weak individual paired with a strong individual at the evolutionary equilibrium, for different population sizes. Average over 30 simulations; $\frac{\beta}{\tau} = 100; \phi = 1$. In the left panel, $x = \frac{1}{2}$. Whiskers represent the min and max value obtained across 30 runs. Additional parameter values can be found in SM section 1.6.

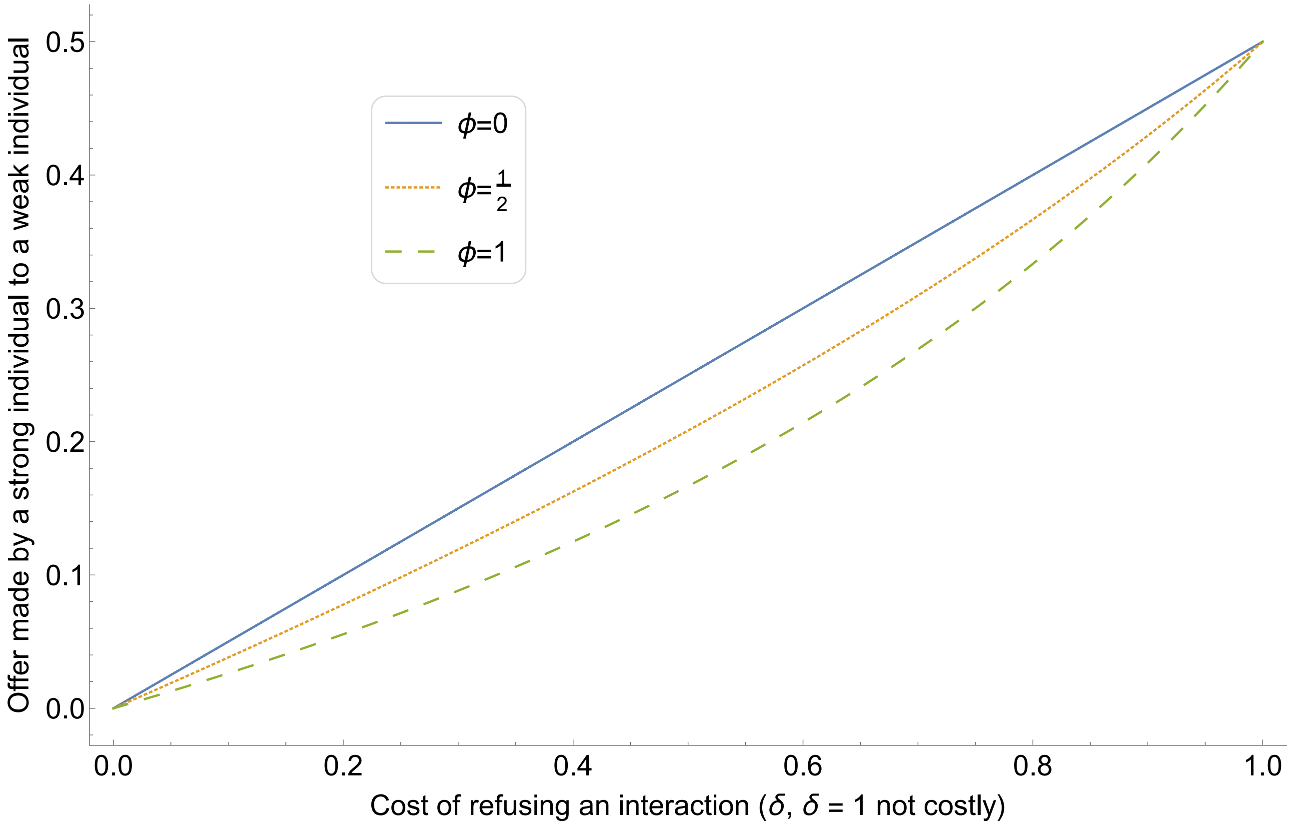

Figure 3: Average offer made by a strong individual to a weak individual at the evolutionary equilibrium, as a function of $\delta$ and for three values of $\phi$ (analytical results). Resource size is 1, $x = \frac{1}{2}$. The degree of dominance of strong individuals $\phi$ has a small impact on the offer they make. When partner choice is not at all costly ($\delta \to 1$), weak individuals receive half of the total resource.

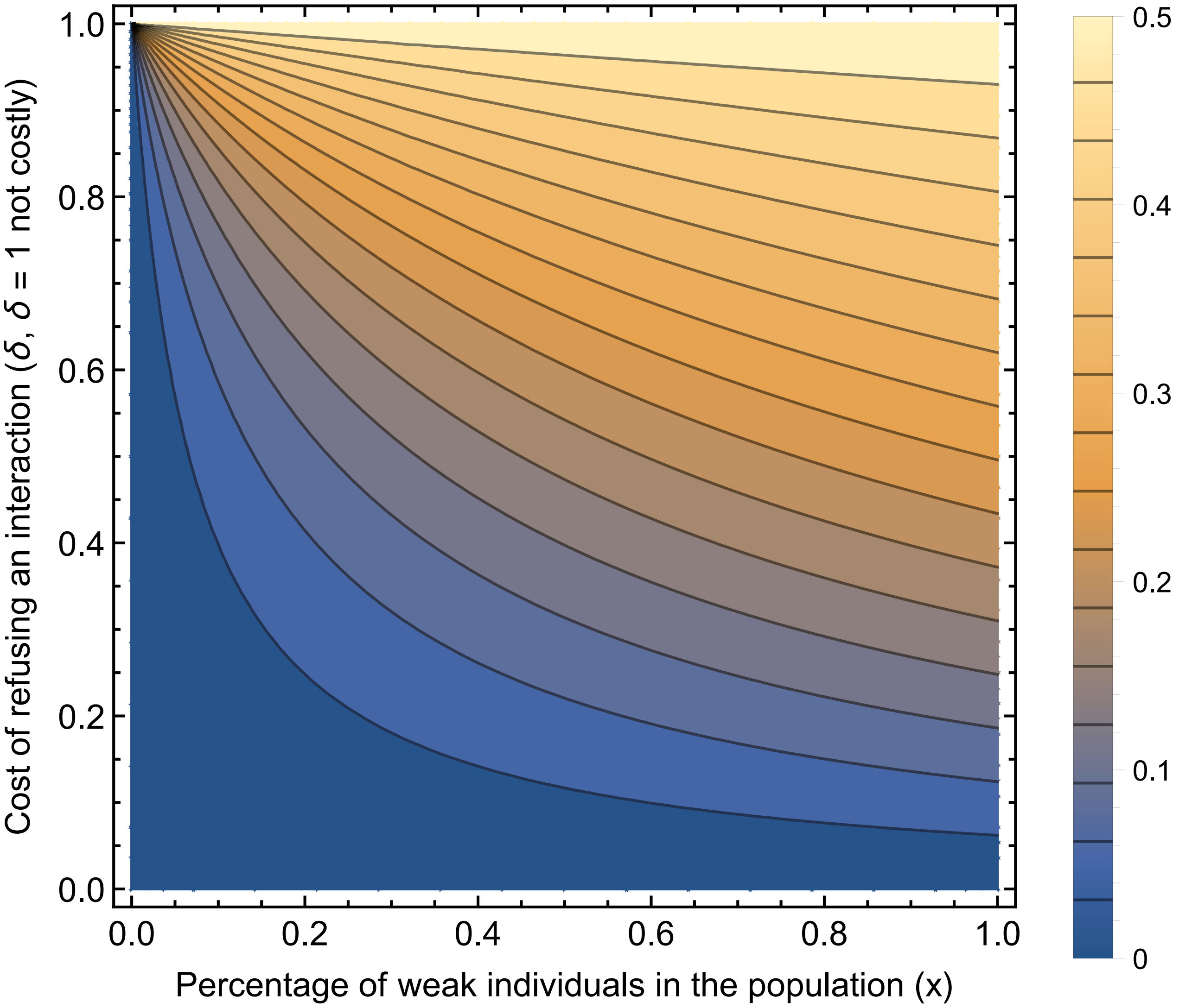

Figure 4: Average offer made by a strong individual to a weak individual at the evolutionary equilibrium as a function of $\delta$ and $x$ (analytical results). Resource size is 1, $\phi = 1$. The percentage of weak individuals in the population has a significant effect on the weak individual's payoff, but this effect is limited when partner choice is not costly.

Supplementary information for paper

# "Evolution of equal division among unequal partners"

**STÉPHANE DEBOVE**[1,2*], **NICOLAS BAUMARD**[3], **JEAN-BAPTISTE ANDRÉ**[1]

[1]Institut de Biologie de l'Ecole normale supérieure (IBENS), INSERM 1024, CNRS 8197, Ecole normale supérieure - PSL Research University, Paris, France.

[2]Université Paris Descartes, Paris, France.

[3]Institut Jean-Nicod (CNRS - EHESS - ENS), Département d'Etudes Cognitives, Ecole normale supérieure - PSL Research University, Paris, France.

[*]**Correspondence**: Stéphane DEBOVE, Institut de Biologie de l'Ecole normale supérieure, 46 rue d'Ulm, 75005 Paris, FRANCE. Email address: debove.stephane@gmail.com

# 1 Supplementary material

## 1.1 Supplementary figures

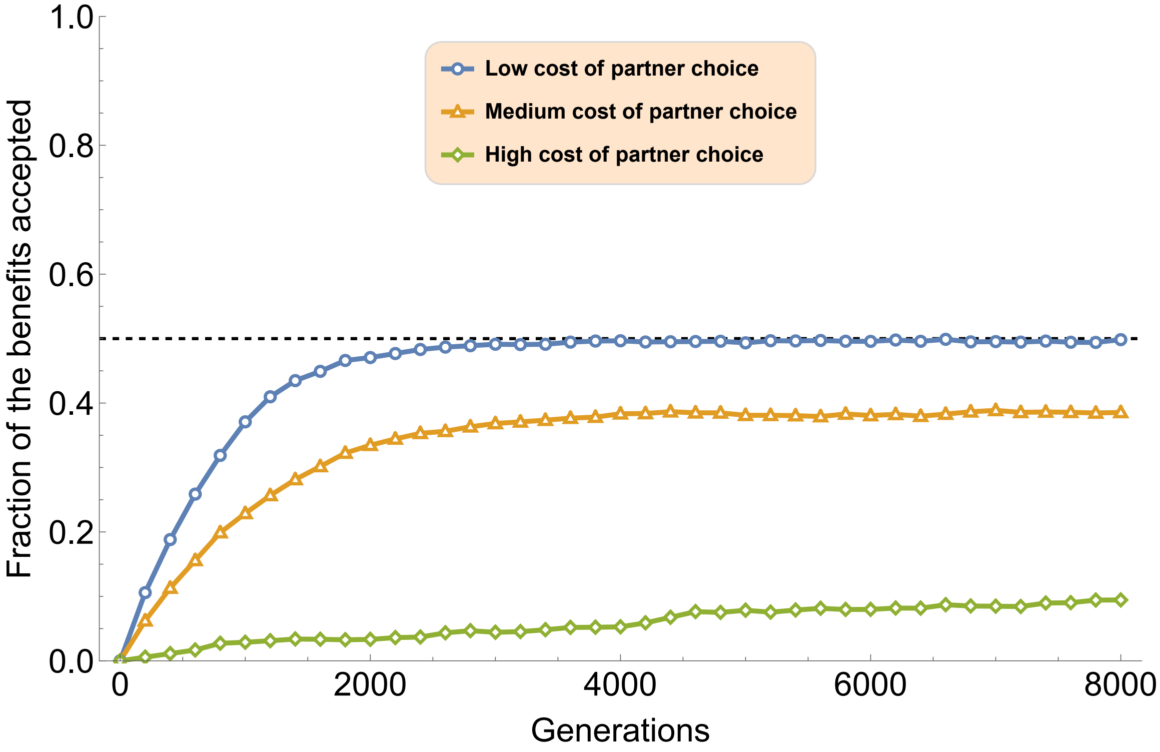

Figure 1: Average offer accepted by a strong individual paired with a strong individual across generations (simulation results). Average over 30 simulations. The dashed line corresponds to the (theoretical) perfectly equal division (50 %). Parameter values: $\phi = 1, x = \frac{1}{2}$. Circle markers: partner choice has almost no cost ($\frac{\beta}{\tau} = 100$). Triangle markers: partner choice has a medium cost ($\frac{\beta}{\tau} = 1$). Diamond markers: partner choice is highly costly ($\frac{\beta}{\tau} = 0.01$). When partner choice is not costly, strong individuals give as much to other strong individuals as they give to weak individuals (see Fig. 1 for comparison). Other parameters used for these simulations can be found in SM section 1.6.

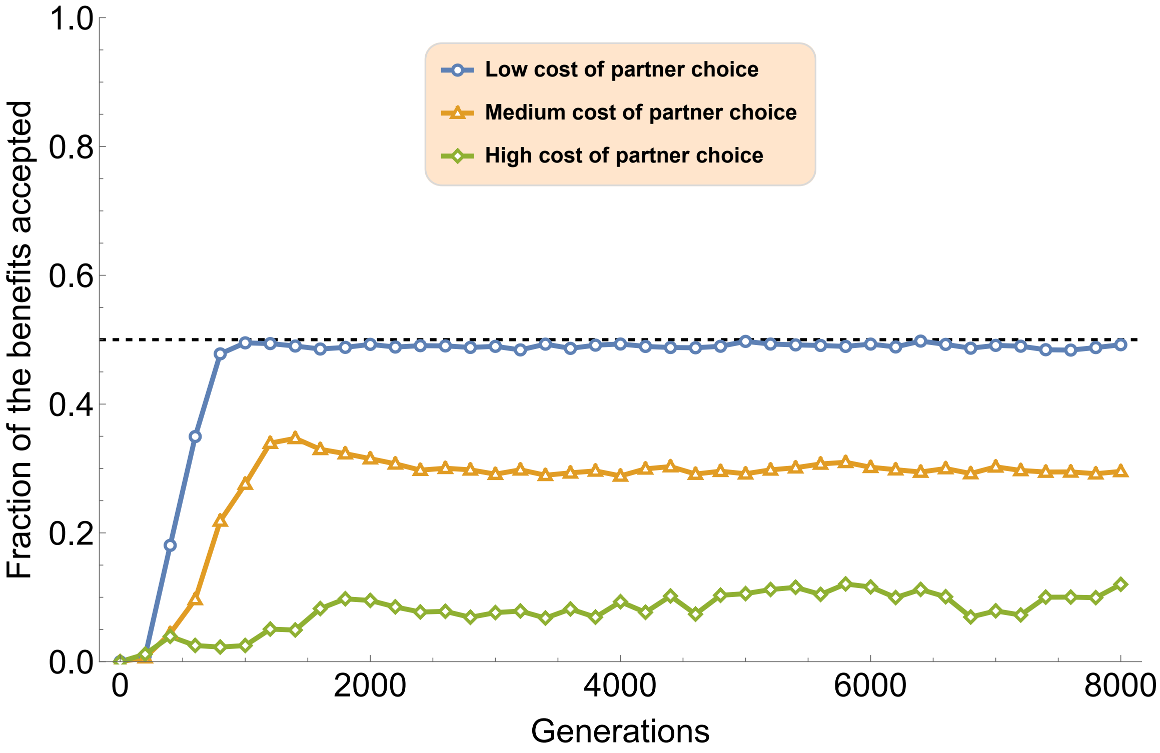

Figure 2: Average benefit received by an individual paired with a stronger individual across generations (continuum of strengths in the population). Average over 30 simulations. Parameter values: $\phi = 1$. Circle markers: partner choice has almost no cost ($\frac{\beta}{\tau} = 100$). Triangle markers: partner choice has a medium cost ($\frac{\beta}{\tau} = 1$). Diamond markers: partner choice is highly costly ($\frac{\beta}{\tau} = 0.01$). When partner choice is not costly, weaker individuals only accept offers that are close to 50% at the equilibrium. Other parameters used for these simulations can be found in SM section 1.6.

## 1.2 Analytical resolution of the model

As mentioned in the main article, the asymmetry of bargaining power between proposers and responders is what makes the UG interesting for studying the evolution of resource division. Because responders are in an inferior bargaining position, in the following analysis we will focus on responders'—and not proposers'—payoffs in the UG. A responder receiving a large share of the resource is a strong indication that there are evolutionary forces at work against the classical proposers' selfishness in the UG.

All our analyses assume that (i) individuals enter the population at a constant rate, (ii) evolution is slow compared to an individual's lifespan (and thus ) (iii) mutations are rare, and that (iv) there is no recombination between genetic traits ($p_{ij}$ and $q_{ij}$). As a consequence of (i) and (ii), the composition of the population does not change during an individual's life. As a consequence of (iii) and (iv), at any evolutionary equilibrium, all the strategies present in the population must reach the same payoff for individuals of a given strength (only a high mutation rate or recombination rate could continuously reintroduce maladaptive strategies in the population, yielding a variance of payoffs at each generation).

Here we make the reasoning used to answer the question posed in the main article explicit: namely, how will the behavioural traits $p_{ij}$ and $q_{ij}$ ($i$ and $j \in \{s, w\}$) evolve in an environment where strong and weak individuals coexist and share resources? As a reminder, $q_{ws}$ reads as "the minimum offer that a weak individual will accept from a strong individual," and $p_{sw}$ as "the offer a strong individual will make to a weak individual."

Following the precise evolutionary dynamics of the system to answer this question is quite a complex challenge, in particular due to epistasis phenomena. The low fitness benefits brought by an offer $p$ can be compensated by high benefits from an acceptance threshold $q$, or small benefits obtained in interactions with individuals of one strength could be compensated by high benefits received in interactions with the other strength, generating linkage disequilibrium (McNamara et al., 2008). But as in (André and Baumard, 2011), it is easier to derive simple conditions on the payoff an individual would or would not have an interest in accepting at the evolutionary equilibrium.

## 1.3 Solving the system

The reasoning is more normative than descriptive, as we consider a situation in which the equilibrium has already been reached, and derive constraints on the values of traits that individuals should display at the equilibrium. To derive the payoff a weak individual should receive from a strong individual at the evolutionary equilibrium, we need to consider four arguments:

1. **All individuals with the same strength must gain the same payoff.** At the equilibrium, all strong individuals should gain the same payoff $G_s$ per interaction (otherwise it wouldn't be an equilibrium), and the same is true for weak individuals. We thus only have two average payoffs in the population at the equilibrium. The average payoff of a strong individual is labeled $G_s$, and that of a weak individual is written $G_w$.

2. **Every individual with strength $i$ accepts exactly $\delta G_i$**, with $i \in \{s, w\}$. If an individual's average payoff is $G_i$, his expected payoff in the next interaction (if the current interaction is refused) will be $\delta G_i$. As a consequence, an individual in the position of responder in the UG should never refuse an offer that is above the corresponding $\delta G_i$, but should always refuse offers that are below this level. At the equilibrium, because offers from proposers should evolve toward the minimum that responders will accept, individuals will always demand and accept exactly $\delta G_i$, no matter who they are interacting with (regardless of their partner's strength). We thus have:

$$\begin{cases} q_{ss} = \delta G_s \\ q_{sw} = \delta G_s \\ q_{ww} = \delta G_w \\ q_{ws} = \delta G_w \end{cases} \tag{1}$$

   Note that this reasoning assumes that acceptance thresholds $q$ are optimized by natural selection, whereas this is not necessarily the case in practice because a certain variability in offers is necessary for this optimization to take place. Nonetheless, André and Baumard (André and Baumard, 2011) showed that even without this assumption, the central results of this modeling approach hold in the simple case where there is only one class of individuals in the population. Here we confirm through simulations that this is also the case when two strengths coexist in the population.

3. **Proposers give responders what they want at the evolutionary equilibrium.**

   Knowing (1) and (2), it can be shown that proposers are always better off giving responders what they "ask for" ($\delta G_i$) at the evolutionary equilibrium, as long as $\delta < 1$. The reasoning is as follows.

   Suppose that at the evolutionary equilibrium, all strong individuals refuse to give weak individuals what they ask for, namely $\delta G_w$ (but all other demands are satisfied). The average social payoff of a strong individual in this population is then

$$G_s = \frac{(1-x)}{2} + x(\frac{1}{2}(1+\phi)\delta G_s + \frac{1}{2}(1-\phi)\delta G_s) \tag{2}$$

As in the main text, $x$ is the proportion of weak individuals in the population, $\phi$ the probability that a strong individual will be the proposer in the UG when paired with a weak individual, and the size of the UG resource to be shared is 1. $G_s$ can be decomposed into three terms: an average payoff obtained in interactions with another strong individual ($\frac{1}{2}$), an average payoff obtained in interactions with a weak individual as proposer (in this case, the offer will be rejected and the strong individual's payoff will be discounted by $\delta$), and, finally, an average payoff obtained in interactions with a weak individual as responder (the strong individual's demand is met, $\delta G_s$).

Similarly, the payoff of a weak individual in this population is

$$G_w = \frac{x}{2} + (1-x)(\frac{1}{2}(1-\phi)(1-\delta G_s) + \frac{1}{2}(1+\phi)\delta G_w) \tag{3}$$

Solving the system composed of equations (2) and (3) gives us an expression for $G_s$ and $G_w$. The question we need to answer now is the following: what would happen if, in such a population, a mutant strong individual decided to accept to give weak individuals what they want? Upon meeting a weak individual and being assigned the role of proposer, this mutant would gain $1-\delta G_w$ (the resource to be shared minus the demand of a weak individual) instead of just $\delta G_s$ (the average payoff being discounted). Knowing $G_w$ and $G_s$, it is easy to show that it is never possible that $\delta G_s \geq 1-\delta G_w$ as long as $\delta < 1$. In other words, at the evolutionary equilibrium, it is impossible that all strong individuals refuse to offer $\delta G_w$ to weak individuals, because they would gain more from doing so.

What if there was some polymorphism in the population such that only *some* strong individuals refuse to give weak individuals what they ask for? The average social payoff of those strong individuals is still written the same as in equation (2). But because we know that at the evolutionary equilibrium all individuals with the same strength must gain the same payoff, the payoff of strong individuals will be the same, regardless of phenotype. The coexistence of two types of strong individuals in the population would imply that $\delta G_s = 1-\delta G_w$ (the payoff of the two types of strong individuals in the position of proposer when paired with weak individuals is equal), but as we showed above, this is not possible as long as $\delta < 1$. As a consequence, it is not only impossible that *all* strong individuals refuse to give weak individuals what they want at the evolutionary equilibrium, it is also impossible that *some* strong individuals refuse to give weak individuals what they want as long as $\delta < 1$.

Following the same reasoning, it can be shown that it is not possible for some individuals (of any strength) to refuse to give their social partner (of any strength) what they ask for at the evolutionary equilibrium (see SM section 1.4), and we thus have:

$$\begin{cases} p_{ss} = \delta G_s \\ p_{sw} = \delta G_w \\ p_{ww} = \delta G_w \\ p_{ws} = \delta G_s \end{cases} \tag{4}$$

4. **No offer in the UG is never refused**

   From step (3) it directly results that no offer in the UG is ever rejected at the evolutionary equilibrium, because each proposer's offer is exactly equal to the responder's acceptance threshold, and thus each offer is accepted. If no offer in the UG is ever refused, the average payoff of strong and weak individuals respectively can be written as:

$$\begin{cases} G_s = \frac{1-x}{2} + x(\frac{1}{2}(1+\phi)(1-\delta G_w) + \frac{1}{2}(1-\phi)\delta G_s) \\ \\ G_w = \frac{x}{2} + (1-x)(\frac{1}{2}(1-\phi)(1-\delta G_s) + \frac{1}{2}(1+\phi)\delta G_w) \end{cases} \tag{5}$$

Solving this system gives us an expression for $G_s$ and $G_w$ as a function of $x$,$\delta$ and $\phi$ at the evolutionary equilibrium:

$$\begin{cases} G_s = \frac{-\delta(\phi+1)+2x\phi+2}{\delta((4x-2)\phi-2)+4} \\ \\ G_w = \frac{\delta(\phi-1)+2(x-1)\phi+2}{\delta((4x-2)\phi-2)+4} \end{cases} \tag{6}$$

Equations (4) and (6) can be used to derive the value of $p_{sw}$ that is plotted in Figs. 3 and 4 of the main paper:

$$p_{sw} = \delta\frac{\delta(\phi-1)+2(x-1)\phi+2}{\delta((4x-2)\phi-2)+4} \tag{7}$$

## 1.4 Verification that proposers are always better off accepting the responder's demand at the evolutionary equilibrium

There are four hypothetical primary situations that need to be taken into account:

- a: when strong individuals are proposers, they refuse to meet the demands of other strong individuals
- b: when strong individuals are proposers, they refuse to meet the demands of weak individuals
- c: when weak individuals are proposers, they refuse to meet the demands of other weak individuals
- d: when weak individuals are proposers, they refuse to meet the demands of strong individuals

These situations are not mutually exclusive, however, so the total number of possible situations is:

$\Sigma_{k=1}^{4}\begin{pmatrix}4\\k\end{pmatrix} = 15$

Situation b was already proven to be impossible at the evolutionary equilibrium in the previous section. We now show that the same holds for the 14 remaining situations, by giving the expected social payoff of strong and weak individuals in each situation. We also give the condition that must be satisfied for each situation to be possible at the evolutionary equilibrium; it is then straightforward to show that, given our parameter values ($0 \leq x \leq 1$, $0 \leq \delta < 1$ , $0 \leq \phi \leq 1$), this condition can never be satisfied.

Situation a:

- $G_w = (1-x)\left(\frac{1}{2}\delta(\phi+1)G_w + \frac{1}{2}(1-\phi)(1-\delta G_s)\right) + \frac{x}{2}$
- $G_s = x\left(\frac{1}{2}(\phi+1)(1-\delta G_w) + \frac{1}{2}\delta(1-\phi)G_s\right) + \delta(1-x)G_s$
- Condition $1-\delta G_s \leq \delta G_s$ impossible

Situation c:

- $G_w = (1-x)\left(\frac{1}{2}\delta(\phi+1)G_w + \frac{1}{2}(1-\phi)(1-\delta G_s)\right) + \delta x G_w$
- $G_s = x\left(\frac{1}{2}(\phi+1)(1-\delta G_w) + \frac{1}{2}\delta(1-\phi)G_s\right) + \frac{1-x}{2}$
- Condition $1-\delta G_w \leq \delta G_w$ impossible

Situation d:

- $G_w = (1-x)\left(\frac{1}{2}\delta(1-\phi)G_w + \frac{1}{2}\delta(\phi+1)G_w\right) + \frac{x}{2}$
- $G_s = x\left(\frac{1}{2}(\phi+1)(1-\delta G_w) + \frac{1}{2}\delta(1-\phi)G_s\right) + \frac{1-x}{2}$
- Condition $1-\delta G_s \leq \delta G_w$ impossible

Situation a & c:

- $G_w = (1-x)\left(\frac{1}{2}\delta(\phi+1)G_w + \frac{1}{2}(1-\phi)(1-\delta G_s)\right) + \delta x G_w$
- $G_s = x\left(\frac{1}{2}(\phi+1)(1-\delta G_w) + \frac{1}{2}\delta(1-\phi)G_s\right) + \delta(1-x)G_s$
- Condition $1-\delta G_w \leq \delta G_w \wedge 1-\delta G_s \leq \delta G_s$ impossible

Situation b & c:

- $G_w = (1-x)\left(\frac{1}{2}\delta(\phi+1)G_w + \frac{1}{2}(1-\phi)(1-\delta G_s)\right) + \delta x G_w$
- $G_s = x\left(\frac{1}{2}\delta(1-\phi)G_s + \frac{1}{2}\delta(\phi+1)G_s\right) + \frac{1-x}{2}$
- Condition $1-\delta G_w \leq \delta G_s \wedge 1-\delta G_w \leq \delta G_w$ impossible

Situation c & d:

- $G_w = (1-x)\left(\frac{1}{2}\delta(1-\phi)G_w + \frac{1}{2}\delta(\phi+1)G_w\right) + \delta x G_w$
- $G_s = x\left(\frac{1}{2}(\phi+1)(1-\delta G_w) + \frac{1}{2}\delta(1-\phi)G_s\right) + \frac{1-x}{2}$
- Condition $1-\delta G_w \leq \delta G_w \wedge 1-\delta G_s \leq \delta G_w$ impossible

Situation b & d:

- $G_w = (1-x)\left(\frac{1}{2}\delta(1-\phi)G_w + \frac{1}{2}\delta(\phi+1)G_w\right) + \frac{x}{2}$
- $G_s = x\left(\frac{1}{2}\delta(1-\phi)G_s + \frac{1}{2}\delta(\phi+1)G_s\right) + \frac{1-x}{2}$
- Condition $1-\delta G_w \leq \delta G_s \wedge 1-\delta G_s \leq \delta G_w$ impossible

Situation a & d:

- $G_w = (1-x)\left(\frac{1}{2}\delta(1-\phi)G_w + \frac{1}{2}\delta(\phi+1)G_w\right) + \frac{x}{2}$
- $G_s = x\left(\frac{1}{2}(\phi+1)(1-\delta G_w) + \frac{1}{2}\delta(1-\phi)G_s\right) + \delta(1-x)G_s$
- Condition $1-\delta G_s \leq \delta G_s \wedge 1-\delta G_s \leq \delta G_w$ impossible

Situation a & b:

- $G_w = (1-x)\left(\frac{1}{2}\delta(\phi+1)G_w + \frac{1}{2}(1-\phi)(1-\delta G_s)\right) + \frac{x}{2}$
- $G_s = x\left(\frac{1}{2}\delta(1-\phi)G_s + \frac{1}{2}\delta(\phi+1)G_s\right) + \delta(1-x)G_s$
- Condition $1-\delta G_w \leq \delta G_s \wedge 1-\delta G_s \leq \delta G_s$ impossible

Situation a & c & d:

- $G_w = (1-x)\left(\frac{1}{2}\delta(1-\phi)G_w + \frac{1}{2}\delta(\phi+1)G_w\right) + \delta x G_w$
- $G_s = x\left(\frac{1}{2}(\phi+1)(1-\delta G_w) + \frac{1}{2}\delta(1-\phi)G_s\right) + \delta(1-x)G_s$
- Condition $1-\delta G_s \leq \delta G_s \wedge 1-\delta G_w \leq \delta G_w \wedge 1-\delta G_s \leq \delta G_w$ impossible

Situation a & b & c:

- $G_w = (1-x)\left(\frac{1}{2}\delta(\phi+1)G_w + \frac{1}{2}(1-\phi)(1-\delta G_s)\right) + \delta x G_w$
- $G_s = x\left(\frac{1}{2}\delta(1-\phi)G_s + \frac{1}{2}\delta(\phi+1)G_s\right) + \delta(1-x)G_s$
- Condition $1-\delta G_w \leq \delta G_w \wedge 1-\delta G_s \leq \delta G_s \wedge 1-\delta G_w \leq \delta G_s$ impossible

Situation b & c & d:

- $G_w = (1-x)\left(\frac{1}{2}\delta(1-\phi)G_w + \frac{1}{2}\delta(\phi+1)G_w\right) + \delta x G_w$
- $G_s = x\left(\frac{1}{2}\delta(1-\phi)G_s + \frac{1}{2}\delta(\phi+1)G_s\right) + \frac{1-x}{2}$
- Condition $1-\delta G_w \leq \delta G_w \wedge 1-\delta G_s \leq \delta G_w \wedge 1-\delta G_w \leq \delta G_s$ impossible

Situation a & b & d:

- $G_w = (1-x)\left(\frac{1}{2}\delta(1-\phi)G_w + \frac{1}{2}\delta(\phi+1)G_w\right) + \frac{x}{2}$
- $G_s = x\left(\frac{1}{2}\delta(1-\phi)G_s + \frac{1}{2}\delta(\phi+1)G_s\right) + \delta(1-x)G_s$
- Condition $1-\delta G_s \leq \delta G_s \wedge 1-\delta G_w \leq \delta G_s \wedge 1-\delta G_s \leq \delta G_w$ impossible

Situation a & b & c & d:

- $G_w = \delta(1-x)G_w + \delta x G_w$
- $G_s = \delta(1-x)G_s + \delta x G_s$
- Condition $1-\delta G_s \leq \delta G_s \wedge 1-\delta G_w \leq \delta G_s \wedge 1-\delta G_w \leq \delta G_w \wedge 1-\delta G_s \leq \delta G_w$ impossible

As explained in the previous section, the verification that it is not possible for *some* (but not all) individuals not to interact with other individuals at the evolutionary equilibrium (in case of polymorphism) is already implied by the use of not strict inequalities.

## 1.5 Verification that results are unchanged when using an absolute cost instead of a discount factor

Suppose that when an individual rejects an offer, his average payoff is not discounted by $\delta$ but is decreased by a fixed amount $c$. Knowing that no offer in the UG is ever refused, we can rewrite the average payoff of strong and weak individuals respectively at the evolutionary equilibrium as :

$$\begin{cases} G_s = \frac{1-x}{2} + x(\frac{1}{2}(1+\phi)(1-(G_w-c)) + \frac{1}{2}(1-\phi)(G_s-c)) \\ \\ G_w = \frac{x}{2} + (1-x)(\frac{1}{2}(1-\phi)(1-(G_s-c)) + \frac{1}{2}(1+\phi)(G_w-c)) \end{cases} \tag{8}$$

Solving this system gives us an expression for $G_s$ and $G_w$ as a function of $x$,$c$ and $\phi$ at the evolutionary equilibrium:

$$\begin{cases} G_s = \frac{\phi((4c+2)x-1)+1}{(4x-2)\phi+2} \\ \\ G_w = \frac{\phi(4c(x-1)+2x-1)+1}{(4x-2)\phi+2} \end{cases} \tag{9}$$

From here it is straightforward to show that both $G_s$ and $G_w$ tend towards 0.5 when $c = 0$ and $\phi = 1$, as it was the case when using a discount factor instead of an absolute cost of changing partner.

## 1.6 Simulation parameters

All simulations were coded in Netlogo (Wilensky, 1999) and ran on a dedicated cluster of computers managed by HTCondor (University of Wisconsin–Madison, 2013). Data analysis and graphical representations (including those from the analytical model) were obtained using Mathematica (Wolfram Research, 2012). The code for all simulations and data analysis is available on request from the first author of this paper.

**Parameters Figure 1 et SM1**

Initial $p_{ij}$: 0
Initial $q_{ij}$: 0
Population size: 500
Lifespan: 500
Generations: 8,000
Mutation rate: 0.002
Mutation standard deviation: 0.02
$\phi$: 1
$x$: 0.5
$\tau$: 0.01
$\beta$: 1, 0.01 or 0.0001

**Parameters Figure 2**

Initial $p_{ij}$: 0
Initial $q_{ij}$: 0
Population size: 500
Lifespan: 500
Generations: 8,000
Mutation rate: Left panel: 0.02, right panel: 1/population size
Mutation standard deviation: Left panel: 0.05, Right panel: 0.02

$\phi$: 1
$x$: Left panel: variable, Right panel: 0.5
$\tau$: 0.01
$\beta$: Left panel: 1, Right panel: 10

**Parameters Figure SM2**

Initial $\gamma$: 0
Initial $\rho_p$: -1
Initial $\rho_r$: 1
Initial $\mu$: 0
Initial $\lambda_p$: -1
Initial $\lambda_r$: -1
Population size: 500
Lifespan: 500
Generations: 8,000
Mutation rate: 0.02
Mutation standard deviation: 0.1
$\phi$: 1
$\tau$: 0.01
$\beta$: 1, 0.01 or 0.0001

# References

Jean-Baptiste André and Nicolas Baumard. Social opportunities and the evolution of fairness. *Journal of theoretical biology*, 289:128–35, November 2011. ISSN 1095-8541. doi: 10.1016/j.jtbi.2011.07.031. URL `http://www.ncbi.nlm.nih.gov/pubmed/21907216`.

JM McNamara, Zoltan Barta, Lutz Fromhage, and AI Houston. The coevolution of choosiness and cooperation. *Nature*, 451(7175):189–192, January 2008. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature06455. URL `http://www.nature.com/doifinder/10.1038/nature06455` `http://www.nature.com/nature/journal/v451/n7175/abs/nature06455.html`.

University of Wisconsin–Madison. HTCondor, 2013.

U. Wilensky. NetLogo, 1999. URL `http://ccl.northwestern.edu/netlogo/`.

Inc. Wolfram Research. Mathematica Edition: Version 9.0, 2012.