Journal of Theoretical Biology 289 (2011) 128–135

Contents lists available at SciVerse ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Social opportunities and the evolution of fairness

Jean-Baptiste André [a,*], Nicolas Baumard [b]

[a] Laboratoire Ecologie et Evolution, UMR 7625, CNRS—Ecole Normale Supérieure, 46 rue d'Ulm, 75005 Paris, France
[b] Philosophy, Politics and Economics Program, University of Pennsylvania, 313 Cohen Hall, 249 South 36th Street, Philadelphia, PA 19104, USA

ARTICLE INFO

Article history:
Received 14 June 2011
Received in revised form
22 July 2011
Accepted 26 July 2011
Available online 5 September 2011

Keywords:
Evolution of cooperation
Partner switching
Outside options
Ultimatum game
Moral psychology

ABSTRACT

We model the evolution of the division of a resource between two individuals, according to a bargaining mechanism akin to the ultimatum game, in which a dominant proposer makes an offer that his partner can only accept or refuse. Individuals are randomly drawn from an infinite population and paired two-by-two. In each pair, a proposer is chosen. The proposer offers a division of resources to his partner. If the offer is accepted it is implemented; otherwise both partners pay a cost and move on to the next social opportunity. When the role that individuals play in each interaction is chosen at random, our analysis shows that each individual receives a fraction corresponding to at least $1/2-c$ of the resource at evolutionary equilibrium, where c represents the cost of postponing the interaction. A quasi-fair division thus evolves as long as c is low. We show that fairness, in this model, is a consequence of the existence of an *outside option* for dominated individuals: namely the possibility of playing on terms more favorable to them in the future if they reject the current interaction. We discuss the interpretation and empirical implications of this result for the case of human behavior.

1. Introduction

Cooperative interactions generate surplus benefits and, beyond understanding qualitatively how natural selection has made their existence among non-kin possible, it is also essential to try and understand *quantitatively* how it has shaped the way these benefits are divided. In particular, although the surplus generated by a cooperative interaction can in principle be distributed in infinitely many ways, human beings systematically express a preference for *fair* divisions. We tend to offer specific quantitative shares to our partners, and expect them to do the same. In symmetric interactions, for instance, we expect equal divisions, and we avoid interacting with people who act unfairly by keeping more. Here, we aim to understand how natural selection shapes the division of benefits in social interactions, and in particular how the preference for divisions of a specific kind, which we call fair, may have evolved.

More precisely, we aim to understand the evolutionary rationale for the simplest version of fairness: the division of a common resource into two equal halves. In this aim, and taking a step further relative to most models on cooperation, we assume the occurrence of a cooperative interaction between two players, generating a surplus of constant size, and we seek to understand how the partners distribute this surplus. To do so, we need to specify a bargaining mechanism for the division. If the two partners have the same bargaining power, it is understandable that fairness evolves, as no one can be forced to accept an unfavorable outcome (Rubinstein, 1982). However, the distribution of resources among humans is typically not the outcome of a mere power struggle. Human beings do favor fair outcomes, even in asymmetric interactions in which one dominant player could in principle take all, or a disproportionate share, of the benefits. This is what we aim to understand.

To this end, we consider a particularly simple and maximally asymmetric negotiation mechanism: the ultimatum game (UG; Güth et al., 1982; Camerer, 2003). In the UG, two individuals share a common resource, but one of them (the so-called *proposer*) benefits from a strategic advantage: he has the power to definitely commit to a certain allocation of the resource with no option to change his mind afterward. The other (called the *responder*) has no option but to either accept the proposed offer or reject it and receive nothing (in which case the proposer also receives nothing). In such a situation, both rationality and natural selection (in simple cases) lead to a maximally unfair outcome: the *proposer* keeps virtually all of the resource. Independently of the amount of time and energy each individual may have contributed to the initial production of the resource, the *proposer* keeps all, because the negotiating power is on his side.

Our aim is to understand how natural selection can lead to fairness in this interaction. In particular, we want to consider the fact that individuals always have *outside options* on a market of social partners. In the ultimatum game *stricto sensu*, each time an

* Corresponding author. Tel.: +33 1 44322341.
E-mail addresses: jeanbaptisteandre@gmail.com (J.-B. André), nbaumard@gmail.com (N. Baumard).

0022-5193/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2011.07.031

individual rejects an offer she definitively compromises an opportunity for social interaction, i.e. individuals have to choose between accepting the very offer they are made, or receive nothing. Yet, in reality, social life is made up of a rich diversity of social opportunities which one can choose to take up, or not (see e.g. Aktipis, 2004). Hence, we believe that it is a mistake to consider any given pairwise interaction as isolated from the relevant outside options. It tends to place an exaggerated importance on purely local power asymmetries.

Recently, inspired in part by the theory of biological markets (Noë et al., 1991; Noë and Hammerstein, 1994, 1995) we have built a model in which proposers make their offer in public and responders can choose the proposer with whom they want to interact before the interaction, thus giving rise to a "social market" (André and Baumard, 2011). Our analysis shows that a fair division of the resource can evolve, provided individuals have the option of choosing (i) their partner, and (ii) the role they wish to play (proposer or responder). One way to understand this result is to realize that the local dominance status of an individual, in a given pairwise interaction, has little influence on the outcome of the interaction if this individual can choose instead to enter into another interaction in which she has a different status.

However, because we initially developed our model to understand the effect of partner choice *per se*, the fact that fairness is fundamentally a consequence of outside options was not easily visible in André and Baumard (2011). Besides, for the sake of simplicity we had to make a number of assumptions. In particular, we assumed that partner choice was perfect and costless. Our aim in the present paper is to develop a further model in which (i) these assumptions are relaxed and (ii) fairness is more clearly shown to be a consequence of outside options in a social market.

To this end, we consider a social interaction based on the ultimatum game. Whereas André and Baumard (2011) considered an idealized paradigm of partner choice, in which responders choose the best among all available offers, in this paper we consider a more parsimonious mechanism based on sequential pairing (see also McNamara et al., 2008). Individuals are randomly drawn from an infinite population and paired two-by-two. In each pair, a proposer is chosen. The proposer offers a resource division to his partner. If the offer is accepted it is implemented. If the offer is rejected, rather than receive a nil payoff, both partners pay a cost (for having postponed their interaction) and move on to the next social opportunity (their "outside option"): they are paired randomly with another partner, and so on.

Importantly, as in Nowak et al. (2000), we assume that the role an individual happens to play in a given interaction is chosen *at random*, i.e. there is no intrinsic property of individuals that is correlated to their probability of being chosen as a proposer/responder. This assumption is meant to represent "social fluidity," i.e. the diversity of social interactions that an individual faces in the course of social life. For the sake of comparison, variations within the same basic model will be considered. In particular, we consider (i) a model in which individuals are stably characterized by a role they play throughout their social life, and (ii) a model in which individuals always remain with the same partner but can change role from one interaction opportunity to the next.

2. General presentation of the model

We consider a simple social interaction based on the ultimatum game (UG). Individuals from an infinitely large population are randomly paired. Each pair of individuals is offered a resource of a given constant size $R=1$ and the opportunity to divide it. One individual in the pair (called the proposer) is strategically dominant, i.e. he is able to propose and commit to a division of the resource, whereas the other (called the responder) has only two options: accept the offer, or reject it (in which case both players receive nothing) and hope for a better social opportunity in the future. Depending on the version of the model, each individual's role is chosen either at random or as a function of intrinsic individual properties. Individuals are genetically characterized by (i) the offer p they make when they play the role of proposer, and (ii) the minimum share of the resource q that they accept when they play the role of responder, called their "acceptance threshold." In a given interaction, the offered split is implemented iff $p \geq q$, otherwise the interaction is canceled (and what occurs next depends on the version of the model).

In the ultimatum game *stricto sensu*, each time an individual rejects an offer she definitively compromises an opportunity for social interaction. This favors undemanding responders, as there is no benefit in rejecting offers. Here, we aim to explore the opposite situation, in which individuals have the option of refusing a social interaction without definitively compromising their chance of interacting later. In other words, we wish to explore the consequences of the existence of a *competition* between various opportunities: rejecting a current opportunity opens up the possibility of accepting another, a form of competition that is absent from the UG. Accordingly, in all versions of the analytical model (but not in individual-based simulations), we assume that the total number of *effective* social interactions each individual undergoes per unit of time is constant (by "effective" we mean an interaction in which the proposer's offer is accepted): it does not depend on the time taken to complete each interaction (e.g. the time it takes to find a compatible partner). Therefore, individuals pay a cost for postponing an interaction in terms of energy consumed and/or in terms of time available for other activities (non-social activities, or other social activities), but not in terms of time available for the very social activity under scrutiny.

In all versions of the analytical model (but not in simulations), we assume that (i) individuals in need of a social interaction enter the population at a constant rate, and that (ii) evolution is slow at the scale of individual lifespan. As a result, the composition of the population of potential partners is considered to be constant across the entire life of an individual.

All analyses assume that mutations are rare, and that recombination is absent (between offer p and acceptance threshold q). As a result, in any evolutionary equilibrium, all the strategies present in the population must reach the same payoff.

The cost of postponing the interaction is measured by a factor $\delta \leq 1$. Consider a social payoff of value 1 obtained immediately. If an offer is rejected and the actual interaction is postponed until the next offer (with the same or a different partner, depending on the version of the model), the very same social payoff will then be worth δ, and δ^2 if the interaction is postponed again, and then δ^3, etc. When $\delta=1$, postponing the interaction is free. When $\delta=0$, postponing the interaction is not an actual option and the game becomes in practice an ultimatum game (refusing an offer is like forgoing any payoff). In practice, our analyses will neglect the situation in which postponing the interaction is completely free ($\delta=1$), as it leads to artifactual neutralities.

Our model assumes that individuals have the option of leaving an interaction without entirely losing the investment they have made in this interaction, once they know (i) which role they play (proposer or responder) and (ii) what fraction of the resource the proposer offers. This is a crucial assumption. It can be interpreted in two different ways. First, the proposer has the ability to commit definitely to a given partition of the resource before the interaction takes place, and the responder then accepts or refuses. Second, the responder has some information on the proposer's usual behavior, either because they actually interact repeatedly (the UG is played several times in a row) or because the proposer