SONDERFORSCHUNGSBEREICH 504

Rationalitätskonzepte,
Entscheidungsverhalten und
ökonomische Modellierung

No. 03-32

Reciprocity - an indirect evolutionary analysis

Siegfried K. Berninghaus*
and Christian Korth**
and Stefan Napel***

October 2003

Financial support from the Deutsche Forschungsgemeinschaft, SFB 504, at the University of Mannheim, is gratefully acknowledged.

*Universität Karlsruhe, email: sbe@vwl3.wiwi.uni-karlsruhe.de

**, email: korth@stanfordalumni.org

***Department of Economics, Boston University, email: napel@bu.edu

Universität Mannheim
L 13,15
68131 Mannheim

Reciprocity – An Indirect Evolutionary Analysis

Siegfried K. Berninghaus
School of Economics
and Business Engineering
University of Karlsruhe
D-76128 Karlsruhe
Germany
Fax: +49-721-608-3491
berninghaus@wiwi.uni-karlsruhe.de

Christian Korth
Niersstrasse 12
D-47533 Kleve
Germany
Fax: +49-2821-583748
korth@stanfordalumni.org

Stefan Napel[1]
University of Hamburg and
Boston University
Department of Economics
Boston, MA 02215
USA
Fax: +1-617-353-4449
napel@bu.edu

Version: September 29, 2003

This paper investigates strategic interaction between rational agents whose preferences evolve over time. Players face a pecuniary 'game of life' comprising the ultimatum game and the dictator game. Utility may but need not be attached to the reciprocation of fair and unfair play by the opponent and equitable payoff distributions as proposed by Falk and Fischbacher (2001). Evolutionary fitness is determined solely by material success – regardless of the motives for its achievement. Agents cannot explicitly condition the social component of their preferences on whether they face the ultimatum or dictator game. Under these conditions, agents develop a strong preference for reciprocation but little interest in an equitable distribution as such. This corresponds to equitable ultimatum offers but full surplus appropriation by dictators. Adding an exogenous constraint on the possible divergence between preference for reciprocation and for an equitable distribution either makes ultimatum divisions asymmetric or dictators become generous depending on the relative frequency of ultimatum and dictator interaction.

[1]Financial support from the German Academic Exchange Service (DAAD) is gratefully acknowledged.

1. INTRODUCTION

In experimental investigations of the *ultimatum game*, participants quite consistently offer 30-50% of an available monetary surplus as first-moving proposers. They reject offers of less than 20% as second-moving responders, which results in no player receiving anything. Particularly the latter observation is hard to reconcile with the assumption that economic actors are rational maximizers of their monetary payoffs.[2] By including a consideration for fairness conventions and reciprocity in players' preferences observations can be explained very well. This is true not only for the ultimatum game, but also many other games for which experimental findings are puzzling from a monetary-payoff maximization point of view.

Neoclassical theory does not restrict preferences to be based only on monetary payoffs or to be strictly monotonic in them. But economic agents who are spiteful, enjoy a warm glow donating money to anonymous strangers, or feel it worthwhile to incur private costs to punish free-riders of public goods have certainly not been the conventional assumption in economics. The award of 2002's Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel to Daniel Kahneman and Vernon Smith for "Foundations of Behavioral and Experimental Economics" is only one indicator that this is changing fast.

In adopting a more realistic view of homo oeconomicus, however, one should be careful to not jump to the other extreme, i..e. to count any observation of supposedly odd behavior as evidence that 'standard assumptions' are wrong and to suppose that human beings universally exhibit the nicer-than-expected character exhibited in some laboratory experiments. The latter would be invalidated by many other experiments, e. g. on market games, in which participants' behavior is explained well by egoistic maximization of monetary rewards. Also, the tale – told in different versions before the advent of behavioral economics – that (an unspecified kind of) evolution would make economic agents behave *as if* they indeed maximized their payoffs in a world of scarce material resources in our view contains some grain of truth. The question is: Under which circumstances is behavior more of the 'standard' homo oeconomicus-type, and which environments defined by which conditions induce human beings to act e. g. like homo reciprocans (cp. Fehr and Gächter 1998) and benevolent dictators?

This paper combines methods of evolutionary and behavioral game theory. We attempt to make progress on the above question about environmental and psychological determinants of material-payoff maximization as opposed to social preferences in a very simple model. We concentrate on a possible preference for reci-

[2] The argument that stakes are simply too small for people to bother about monetary payoff maximization has been countered by experiments in which stakes were equivalent of several month's income of participants (e. g. the experiments conducted in Indonesia by Cameron 1999 or in the Slovak Republic by Slonim and Roth 1998).

procity as formalized by Falk and Fischbacher (2001), in an environment consisting not of just one, but two distinct distribution tasks. As a first approximation of the complexities of the real 'game of life', we study evolution of reciprocity-based preferences in a world where agents randomly face either the ultimatum or the dictator game.[3]

The remainder of the paper is organized as follows: The following section clarifies the terms *reciprocity* and *fairness*, compares recent approaches to the definition of preferences that result in reciprocal behavior, and explains our choice of the Falk-Fischbacher formulation. Section 3 gives a brief introduction to the indirect evolutionary approach, section 4 presents our model, which is analyzed in section 5 with focus on the question: Under which circumstances is a reciprocity-based notion of fairness evolutionary stable? Section 6 compares results to those obtained in a related model by Güth and Napel (2002). Section 7 concludes.

2. RECIPROCITY

Positive (negative) reciprocity refers to the impulse or desire to be kind (unkind) to those who have been kind (unkind) to us.[4] Reciprocity is to be distinguished from simple *altruism*, i. e. unconditional generosity. While the narrow self-interest hypothesis in 'standard theory' fails to explain stylized facts of many experiments, the notion of reciprocity preference sheds a fairly consistent light on possible motivational forces behind a number of observations. This section briefly points to major contributions to the literature which incorporate reciprocal behavior into standard game theory, starting with a stylized description of the experimental observations that motivated them.[5]

2.1. Evidence from Experiments

In experimental *ultimatum games*, most proposers offer between thirty and fifty percent of the available surplus ('pie') to the receiver while offers below twenty percent are rare.[6] Small offers are usually rejected by responders, in which case both players receive a zero share; the rejection rate decreases in the amount offered to the responder. An important observation is that responders more likely accept

[3]What we in the following refer to as a *multi-game environment* can also be viewed as a recurrent version of a comprehensive *single* game, in which Nature has the first move and selects either an ultimatum or dictator subgame.

[4]See Fehr and Gächter (1998) for a more detailed account.

[5]Compare Klein (2000) for a detailed comparison of literature on reciprocity. Also see the recent overview by Fehr and Schmidt (2002).

[6]Analysis of the ultimatum game was pioneered by Güth, Schmittberger, and Schwarze (1982). For surveys on ultimatum games see Thaler (1988), Güth and Tietz (1990), Camerer and Thaler (1995) and Roth (1995).

lower offers when these are known to be made by a random device rather than a human player. In the related *best-shot game*, the proposer can only choose between two payoff distributions – one very advantageous, one very disadvantageous to himself or herself. Here, participants usually accept a higher degree of inequity than in the ultimatum game.[7] In the *dictator game* the second player is no longer responding but must accept any offer by the first player. Proposers usually offer less than in ultimatum games. However, about 80 percent offer a positive amount.[8]

In the *gift-exchange game* the first mover (firm) offers a wage w to the second mover (worker). The worker responds by choosing a level of privately costly effort. Higher wages yield lower payoff for the firms and higher ones for workers, while higher effort levels have the opposite effect. The standard game-theoretic prediction based on individual maximization of payoffs is that workers invariably choose the lowest possible effort level. In anticipation of this, firms will only make the lowest possible wage offer. In contrast to these predictions, many experimental subjects as well as real firms offer wages above the market rate, and are rewarded by positive effort levels increasing in wage.[9]

In *public goods games* agents simultaneously decide whether to pay for the provision of a public good. Experiments show that the amount subjects contribute is increasing in their expectation about the contributions of others.[10] Fehr and Gächter (2000) provide evidence that a significant fraction of subjects is willing to bear additional private costs if detected free-riders can be punished.

All these experimental results strongly suggest that subjects care about more than their own material payoff. Fairness considerations of one sort or the other seem important; analytical models with predictive value should take this into account. Different approaches to this are conceivable.

2.2. Equity-based Approaches

Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) are the most prominent models of agents with preferences that exhibit *inequity aversion*. Agents may increase their utility by sacrificing their own material payoff if by doing so their payoff is closer to their counterparts' payoffs. Preferences can depend on the entire *payoff distribution*, but not on any *intentions* ascribed to other players.

In Fehr and Schmidt (1999), agents dislike inequitable outcomes, but less so

---

[7]The best-shot game was introduced by Harrison and Hirshleifer (1989) and Prasnikar and Roth (1992). See also Falk, Fehr, and Fischbacher (2000).

[8]See, for example, Forsythe, Horowitz, Savin, and Sefton (1994) and Andreoni and Miller (2002).

[9]See, for example, Berg, Dickhaut, and McCabe (1995), Falk and Gächter (2002) or Fehr, Gächter, and Kirchsteiger (1996).

[10]For surveys on public goods games see Ledyard (1995) and Daws and Thaler (1988).

when these are to their own advantage. Despite its simplicity, many stylized facts can be explained by this assumption. The model is consistent with giving in dictator, gift-exchange and closely related trust games and with the rejection of low offers in ultimatum games. Since the model does not account for intentions, it fails to explain why people behave differently when playing against a random device instead of a real player, or why low offers in a best-shot game are more readily accepted than in an ultimatum game.

The approach by Bolton and Ockenfels (2000) is similar. Their model contains fewer assumptions about functional forms than that of Fehr and Schmidt and agents compare own material payoff to the group's average, making the reference point for fairness considerations endogenous. Marginal disutility of small deviations from equality is assumed to be zero; so if subjects are non-satiated in their own material payoff they will never propose an equal split in the dictator game. The model can explain giving in dictator and gift-exchange games as well as rejections in ultimatum games. However, it fails to explain punishment patterns in public goods games. Neither can it explain behavior that depends on inequities between other players.[11]

2.3. Intention-based Approaches

A crucial feature of the psychology of reciprocity is that decisions to be kind or unkind to others are based not only on *material consequences* implied by other players' actions but also on the *intentions* attributed to these players. Agents who are motivated by reciprocity discriminate between players who take an (un)generous action by choice and those who are forced to do so.[12]

Prominent formal formalizations of reciprocity in terms of intentions have been given by Rabin (1993), Dufwenberg and Kirchsteiger (2001) and Falk and Fischbacher (2001). Rabin was the first to adopt the framework of psychological games of Geanakoplos, Pearce, and Stacchetti (1989) to model reciprocity. He introduced so-called *fairness games* in which a reciprocity payoff is added to the material payoff of the players. The reciprocity payoff is calculated as the product of a *kindness term* and a *reciprocity term*. The kindness term is positive whenever a player feels

[11]For example, Charness and Rabin (2002) let a player C choose between the material payoff allocations $(x_A, x_B, x_C) = (575, 575, 575)$ and (900,300,600), in both of which player C receives his or her fair share of $\frac{1}{3}$. Bolton and Ockenfels' model predicts the second choice as the payoff for player C is higher, but 54% of subjects choose the first allocation in the experiments.

[12]Experimental evidence is given in Falk, Fehr, and Fischbacher (2003) for four different mini-ultimatum games. In each game the proposer had two choices, one of which was always to offer 20%. The alternatives were 0%, 20% 50% or 80% respectively. The rejection rate of the 20% offer was highest when the alternative was equal division, it was lowest when the only alternative was to offer nothing to the second player. The fact that the rejection rate was not zero in this case suggests that pure equity considerations also play a role.

being treated well. Then he or she tries to make the reciprocity term positive, too, in order to increase his or her total utility payoff. This can be achieved by being nice in return, as the reciprocity term is defined to be positive when the player chooses a kind action. Negative reciprocity is modelled analogously.

Rabin's original model is restricted to two-player normal form games. Dufwenberg and Kirchsteiger (2001) generalize Rabin's model to n-person extensive form games of imperfect information. Requiring that behavior is sequentially rational, the main complication is to keep track of the beliefs which determine the attributed intentions as the game evolves.

Falk and Fischbacher (2001) extend Rabin's approach to perfect information extensive form games with finitely many stages. They construct a utility function that allows for pure equity concern *and* for intention-based reciprocity. This combination implies an additional degree of freedom by which – as is to be expected – more experimental observations can be explained than by either pure approach. Equilibrium calculations and parameter calibrations are already complex for entirely intention-based models of reciprocity;[13] this becomes yet more pressing with Falk and Fischbacher's hybrid model.

Both Falk and Fischbacher (2001) and Dufwenberg and Kirchsteiger (2001) produce predictions more or less in line with the experimental observations mentioned above. For the ultimatum game, Dufwenberg and Kirchsteiger's model unfortunately yields multiple sequential reciprocity equilibria (SRE). In all SREs high offers are accepted and low offers are rejected. However, there is no unique prediction for responder reaction to medium-range offers. For every parameter constellation there exists an SRE in which the proposer offers the smallest amount which is acceptable to the responder. For high reciprocity parameters there also exist SREs in which equilibrium offers are rejected, because the players view one another as unkind, which in equilibrium, in fact, has both players act unkindly.

Falk and Fischbacher's reciprocity equilibrium for the ultimatum game involves a mixed responder strategy. Acceptance probability is increasing in the offered amount up to a cut-off level, from which on the offer is always accepted. The cut-off is decreasing in the responder's reciprocity parameter and always lower than one half. The optimal offer by a selfish proposer matches this level. If the proposer is reciprocal, too, his or her offer can but need not be more generous, depending on how reciprocal he or she is relative to the responder. The equilibrium in random-offer ultimatum games, in which the offer is not determined by a human but by a random device, highlights the role of intentions: Because no intentions are attributed to a random offer, the acceptance probability of a given offer is higher than in the regular ultimatum game. It is not necessarily one, because pure equity considerations also play a role.

[13] See Klein (2000, p. 24) for similar thoughts on reciprocity-models.

The model of Falk and Fischbacher (2001) correctly predicts the main stylized facts for the gift-exchange game, too. Worker's effort choice is increasing in wage paid and her reciprocity parameter. Strictly positive wages result. Predictions by the Falk and Fischbacher model for the best-shot game, the dictator game and the public goods game are similarly consistent with the experimental evidence discussed above. This and its comprehensiveness – covering pure inequity aversion or pure intentional reciprocity as special cases – makes it a very powerful tool for the analysis of many different games and our choice in the following analysis.

3. INDIRECT EVOLUTION

The *indirect evolutionary approach* studies the *evolution of preferences* among rational decision makers. Agents are rational in the sense that they have preferences and select strategies with the goal to maximize expected utility. Preferences can but need not depend on an agent's individual material payoff alone, they may e. g. exhibit reciprocity in the way discussed above. However, it is assumed that (average) material payoffs ultimately determine the fitness of agents having particular preferences – and hence preferences themselves – in an evolutionary process: Agents with preferences that earn material payoffs above (below) the average reproduce more (less) successfully, and their and their preferences' population share increase (decrease).

The considered agents make choices based on anticipated consequences, evaluating their options by preferences which evolved dependent on past successes. Thus, as argued in more detail e. g. in Berninghaus, Güth, and Kliemt (2003) and Güth, Kliemt, and Napel (2003), the indirect evolutionary approach combines the main elements of the traditional neoclassical approach (rational purposeful actions selected according to their *anticipated consequences* or the '*shadow of the future*') and the direct evolutionary approach (agents carry out fixed behavioral programs which evolve based only on *past success* or the '*shadow of the past*').[14]

Utility and *material payoff* are distinguished. Agents' objectives – including possibly a preference for reciprocity – are described by their *utility functions*, which under additional conditions imply a unique equilibrium outcome for a given game. Equilibrium material payoffs are interpreted as fitness of the respective

[14]The difference between the direct and indirect evolutionary approaches is smaller than it appears: Models that study the evolution of behavior can be regarded as the special case of preference evolution where feasible preferences are restricted to make a distinct feasible strategy strictly dominant. Models that study preference evolution in a game Γ can be regarded as studying direct evolution in a higher-level game Γ'. The latter can be constructed by identifying each player's strategy set S' with the possible preferences on outcomes of Γ and by taking the payoff function π' to be the composition $\pi \circ \mu$ of the mapping μ from strategy profiles S'^N in Γ' (i. e. a preference regarding outcomes of Γ for each player $i \in N$) to equilibrium strategy profiles of Γ and the original payoff function π.

preference type. A number of studies has used this method to evaluate the stability of different kinds of preferences in specific single-game environments. Huck and Oechssler (1999), for example, explore spiteful preferences in a mini-version of the ultimatum game. Koçkesen, Ok, and Sethi (2000a) and Koçkesen, Ok, and Sethi (2000b) analyze interdependent preferences in symmetric aggregative games and certain classes of symmetric supermodular and submodular games, respectively. Sethi and Somanathan (2001) consider reciprocal preferences in aggregative games. Berninghaus, Güth, and Kliemt (2003) study the trust game and Poulsen and Poulsen (2003) investigate reciprocity, altruism, and materialism in prisoner's dilemma games.

Using the same utility function in different games, it is possible to calculate material success of *general* aspects of preferences in a complex environment. A particular sense of fairness (or absence thereof) seems to be such a general character trait rather than a game-specific preference. The indirect evolutionary approach extended to multiple games thus lends itself to stability tests of the preferences discussed in section 2.

The assumption of preferences for reciprocity has been justified above in terms of its success in making sense of and predicting behavior in experiments. Finding preferences so that they fit a given set of experimental results is useful in a descriptive sense. This may, however, be regarded as merely a question of sufficiently many free parameters and econometric exercise. We want to investigate if the assumption could be justified even without the data, just by answering the question: Which preferences within a given class – which is a restriction already – would one expect to prevail given the axiom that evolutionary success is monotonic in material payoffs? Evolutionary stability of reciprocity would in our view constitute an additional reason to update the common narrow view of homo oeconomicus as a purely self-interested materialist.

The following analysis is motivated by Güth and Napel (2002). There, equity-based preferences similar to those of Fehr and Schmidt (1999) are subjected to evolutionary pressures in an environment consisting of a highly stylized *'game of life'*. It does clearly not reflect the complexities of the real 'game of life' which has shaped the preferences of experimental subjects. However, it seems already a little bit closer to real social environments than any even more stylized single-game environment. Here, the multi-game indirect evolutionary analysis of Güth and Napel[15] is extended to the domain of psychological games, using preferences of the Falk-Fischbacher type, which reflects not only inequity aversion but also intention-based reciprocity. We deem it interesting to see whether the way in which fairness preferences are modelled matters. A direct comparison with Güth and Napel's results is contained in section 6.

[15]Cf. Stahl and Haruvy (2002) for an experimental study of multi-game environments.

4. THE MODEL

Repeatedly, two agents who are randomly drawn from a *single population* are given the chance to create a surplus (the 'pie'), and subsequently to decide about its distribution in a single-shot game. Taking the population to be large enough to rule out repeated-game effects, agents are assumed to act fully rationally according to commonly known preferences for the possible outcomes of the single-shot game presented to them. Utility is not restricted to own material payoff, but it is the expected material payoff which defines an agent's *fitness* or *reproductive success* in the evolutionary process imposed on the population.[16] As expressed in a pointed way by Samuelson (2001, p. 226f): "Nature can thus mislead her agents, in that preferences and fitness can diverge, but cannot mislead herself, in that high fitness wins the day."

4.1. The Material World

A matched pair of agents faces either of two different games: The *ultimatum game* or the *dictator game*. In the *ultimatum game* one of the two players is randomly (with probability 0.5) selected to be the proposer (role X), who proposes how to split the pie of one unit. The amount he or she offers is denoted by $c \in [0, 1]$. The other player (the responder, role Y) decides whether to accept the proposed split or to reject it. The resulting material payoffs are $(\pi_X, \pi_Y) = (1 - c, c)$ and $(\pi_X, \pi_Y) = (0, 0)$, respectively.

In the *dictator game* the agents are also assigned to be proposer or responder with equal probability. The proposer decides how to split the pie by offering $c \in [0, 1]$ to the responder. In contrast to the ultimatum game the responder *must* accept, so resulting material payoffs are $(\pi_X, \pi_Y) = (1 - c, c)$.

Which game the agents play is determined randomly, with probability $\lambda \in [0, 1]$ for the ultimatum game and $1 - \lambda$ for the dictator game. The game realization becomes *common knowledge*. In both events material payoffs (π_X, π_Y) determine reproductive success.

Differences between dictator game versus ultimatum game very loosely resemble those between competitive market interaction versus private interaction: Success in the ultimatum game depends on mutual cooperation and is influenced by both agents' behavior; in the dictator game one player is a mere price taker without influence on terms of trade. It will be interesting to see how evolutionary stable preferences depend on the *frequency* or 'importance' of each type of interaction in agents' lives.

[16]See Benaim and Weibull (2003) for a concise overview of models with agents who learn or imitate rather than biologically reproduce which can be very closely approximated by evolutionary models coming from a purely biological background.

4.2. Agents' Preferences

Agents have *fairness preferences* as defined by Falk and Fischbacher (2001). We prefer not to explicitly restate their quite complex utility function here, since we will directly work with the reciprocity equilibria calculated for dictator and ultimatum games by Falk an Fischbacher. Importantly, utility functions $u_i^X(\cdot)$ and $u_i^Y(\cdot)$ which represent preferences of a given agent i in the roles of proposer X or responder Y, respectively, have the free parameters $\rho_i^X, \varepsilon_i^X$ and $\rho_i^Y, \varepsilon_i^Y$.[17]

Agents' behavior depends on these free parameters. Most important is the *reciprocity parameter* $\rho_i^k \in \mathbb{R}_+$, which describes how much weight agent i places on *reciprocal behavior* in role $k \in \{X, Y\}$. For $\rho_i^k = 0$ the considered agent is purely interested in his or her own *material payoff*, behaving as in most orthodox economic models. The *rejection probability* of ultimatum game offers below 50% turns out to be strictly increasing in reciprocity parameter ρ_i^Y.

The second parameter, $\varepsilon_i^k \in [0, 1]$, measures agent i's pure concern for an equitable outcome in role k. Agents without any pure equity concern ($\varepsilon_i^X = \varepsilon_i^Y = \varepsilon_i = 0$) have an entirely intentions-driven notion of fairness. If no intentions can be attributed to the opponent player, the considered agent maximizes material payoff. In contrast, for agents with $\varepsilon_i > 0$ equitable splits of the pie are valuable even in the absence of intentions. This is best seen in the dictator game where (anticipated or own) reciprocal behavior plays no role: An agent with $\varepsilon_i = 0$ always offers $c = 0$ whereas an agent with $\varepsilon_i = 1$ may offer up to half of the pie.[18]

In principle, preferences could evolve *role-dependently* ($\rho_i^X \neq \rho_i^Y$ and $\varepsilon_i^X \neq \varepsilon_i^Y$), so that an agent's way of evaluating intentions and material consequences of choices depends on his or her role in the game. However, the beauty of the idea to outfit agents with a utility function which reflects *general* agent characteristics like equity concern or reciprocity preference is that agents *generally* apply this utility function to the situations they face – regardless of the actual situation and regardless of their *role* in the situation.[19] Therefore the parameters are assumed to evolve *role-independently* in the following, i. e. $\rho_i^X \equiv \rho_i^Y \equiv \rho_i$ and $\varepsilon_i^X \equiv \varepsilon_i^Y \equiv \varepsilon_i$ for any given

[17] Player i's utility at any terminal node t of the game tree is the sum of his or her *material payoff* associated with t and a *reciprocity payoff* which is calculated decision node by decision node along the path to t and weighted by parameter ρ_i. For each of i's decision nodes n along the path to t, the products of a term measuring the (un)kindness of others to i which has lead to n (here, parameter ε_i may matter) and a term measuring i's reciprocated (un)kindness from choosing to move further towards t are added up. Equilibrium calculations are complex.

[18] The actual offer depends on the pure equity-concern *and* the reciprocity parameters.

[19] The possibility that agents can condition equity concern on the game and / or their role in it is explored in Güth and Napel (2002). If real economic agents' fair behavior is adequately described by a utility function with one or a few fairness-related parameters at all, the same parameters should be valid for more than a special class of games. To expect them to be valid in *all* games seems, however, too much.

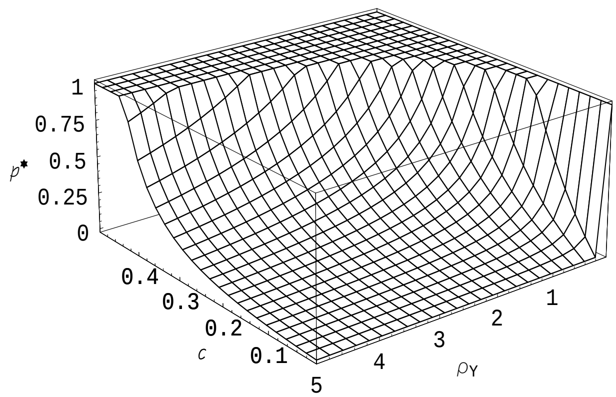

FIG. 1 Acceptance probability p^* of agent with reciprocity parameter ρ_Y for offer c in the ultimatum game

agent. Note that, moreover, the same two parameters ρ_i and ε_i will be applied to both ultimatum and dictator games, i. e. the social part of agent i's preferences is assumed to *not* depend on the realized type of game.

5. EVOLUTIONARY ANALYSIS

First, ultimatum and the dictator games are studied separately, corresponding to the cases $\lambda = 1$ and $\lambda = 0$. Then, the mixed environment consisting of both games is analyzed. When agents A and B interact, we will in the following write ρ_X (ρ_Y) for that agent's (role-independent) reciprocity parameter who is assigned to role X (role Y).

5.1. The Ultimatum Game

Equilibrium play[20] results in acceptance probability

$$p^* = \begin{cases} \min\left\{1, \frac{c}{\rho_Y \cdot (1-2c)(1-c)}\right\} & \text{if} \quad c < \frac{1}{2} \\ 1 & \text{if} \quad c \geq \frac{1}{2} \end{cases} \tag{1}$$

for offer c conditional on the responder's reciprocity parameter $\rho_Y \neq 0$, and $p^* = 1$ for $\rho_Y = 0$. An offer of half of the pie or more is always accepted for sure. This is not the case for lower offers, for which the *acceptance probability* is strictly decreasing in ρ_Y (see figure 1).

The share c^* offered by the proposer X to the responder Y in equilibrium is

[20]The computations are given in detail in Falk and Fischbacher (2001, appendix 3). Cf. footnote 17 above.

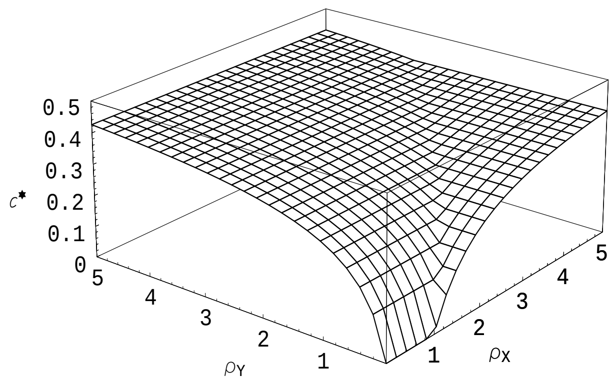

FIG. 2 Split c^* proposed by an agent with reciprocity parameter ρ_X to one with reciprocity parameter ρ_Y in the ultimatum game

given by

$$c^* = \max\left\{\frac{1 + 3\rho_Y - \sqrt{1 + 6\rho_Y + \rho_Y^2}}{4\rho_Y}, \frac{1}{2} \cdot (1 - \frac{1}{\rho_X})\right\} \tag{2}$$

for $\rho_X, \rho_Y \neq 0$. For $\rho_X > \rho_Y = 0$ and $\rho_Y > \rho_X = 0$, c^* is (2)'s limit for $\rho_Y \downarrow 0$ and $\rho_X \downarrow 0$, respectively, while for $\rho_X = \rho_Y = 0$ one obtains $c^* = 0$ (corresponding to 'traditional' preferences).

The equilibrium share c^* is defined by two expressions. The first depends only on the responder's reciprocal inclination ρ_Y and is the maximizer of proposer's expected utility in case that the responder's concern for reciprocity is 'binding' – dominating the insufficiently pronounced reciprocity concern of the proposer. The associated share of the pie is just high enough to ensure acceptance ($p = 1$). The second expression depends only on the *proposer's* ρ_X and reflects how much the proposer would *voluntarily* offer in order to maximize utility in the light of his or her intrinsic concern for a fair outcome. As either ρ_X or ρ_Y grow large, c^* approaches 1/2 (see figure 2).

The equilibrium offer is the maximum of both expressions. This means that if a selfish proposer plays against a reciprocal responder, the offer is increasing in ρ_Y. If the responder's concern for reciprocity is low, the offer depends on the fairness concern of the proposer, hence it is increasing in ρ_X. If both players are selfish, the offer is close to zero.

Agent A's role in interaction with agent B is random, with probability 0.5 for either role. The expected material payoff $\pi^*_{A_U}$ of agent A with preference parameters (ρ_A, ε_A) who is matched with agent B to play the *ultimatum game* is the average of the payoffs he or she receives in both roles. This turns out to be:

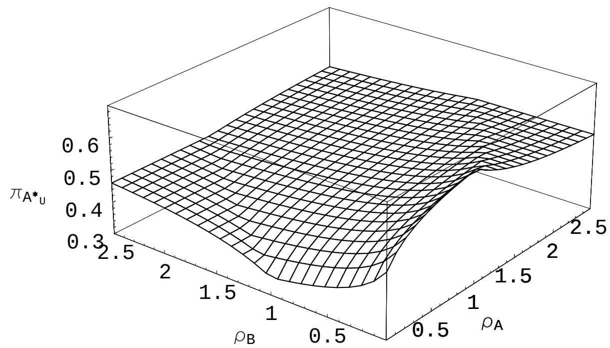

FIG. 3 Expected payoff $\pi^*_{A_U}$ of agent A with reciprocity parameter ρ_A in the ultimatum game when matched against an agent B with reciprocity parameter ρ_B

$$\pi^*_{A_U}(\rho_A,\rho_B) = \begin{cases} \frac{5\rho_A+3-\sqrt{1+6\rho_A+\rho_A^2}}{8\rho_A} & \text{if} \quad \rho_A \geq \frac{-2\rho_B}{\rho_B+1-\sqrt{1+6\rho_B+\rho_B^2}} & \text{(a)} \\ \frac{3\rho_B-3+\sqrt{1+6\rho_B+\rho_B^2}}{8\rho_B} & \text{if} \quad \rho_A \leq \frac{\rho_B(\rho_B-1)}{1+\rho_B} & \text{(b)} \\ \frac{3\rho_A+1-\sqrt{1+6\rho_A+\rho_A^2}}{8\rho_A} & & \\ \quad + \ \frac{\rho_B-1+\sqrt{1+6\rho_B+\rho_B^2}}{8\rho_B} & \text{if} \quad \frac{-2\rho_B}{\rho_B+1-\sqrt{1+6\rho_B+\rho_B^2}} > \rho_A > \frac{\rho_B(\rho_B-1)}{1+\rho_B}. & \text{(c)} \end{cases} \tag{3}$$

Figure 3 illustrates this. Expected payoff depends on both agents' concerns for reciprocity, captured by ρ_A and ρ_B.

There are three cases. In case (a), ρ_A is large compared to ρ_B and thus behavior in *both* possible role realizations is determined by agent A's strong concern for fairness. Agent A's payoff is strictly decreasing in ρ_A. So evolution would drive ρ down until case (c) applies: The population share of agents with payoffs below average (with a higher ρ) will decrease and the share of agents with a payoff above average (with a lower ρ) will increase.[21]

Case (b) applies if ρ_A is small compared to ρ_B. Then, behavior in both roles is determined by ρ_B, and agent A's payoff is constant in ρ_A. Given the symmetry of our setup – both agents are drawn from the *same* population – this is similar to

[21]We are intentionally a bit sloppy in describing evolutionary dynamics. In order to be precise we would have to define a payoff-monotonic selection dynamic which is either started from an interior initial state (in which every ρ is present) or is augmented by an innovative mutation process which ensures that every ρ gets a chance to spread. We refrain from a formal treatment because the technicalities of properly dealing with the state space $\Delta(\mathbb{R}_+)$ consisting of probability distributions μ on the nonnegative real line are immense and do not add anything to our main argument. See Oechssler and Riedel (2002) for a flavor of the mathematical techniques that would be necessary.

case (a), but now from agent B's point of view: Payoff is higher for lower ρ_B so that evolution again drives ρ down until case (c) applies.

In the intermediate case (c), ρ_A and ρ_B do not differ too much. The proposer (regardless of whether agent A or B is given this role) offers a split that maximizes his or her own material payoff subject to the binding constraint imposed by the rejection behavior of the responder, i. e. c^* depends only on the responder's reciprocity parameter. Then agent A's (B's) payoff is increasing in ρ_A (ρ_B). Evolution results in ρ in the entire population approaching infinity within the variability bound defining case (c). This is in the following referred to as ρ 'diverging'.

Whatever the initial distribution of the reciprocity parameter in the population is, ρ will diverge. For the ultimatum game this implies an *equal split* in the limit. It is in the nature of this split, that it is always offered for *strategic reasons*, i. e. to prevent the responder from rejecting, not because of the proposer's intrinsic motivation.

5.2. The Dictator Game

The split c^* offered by the proposer X in the dictator game in equilibrium depends only on his or her *own* reciprocity parameter ρ_X and pure outcome-concern parameter ε_X:

$$c^* = \max\Big\{0, \frac{1}{2} \cdot (1 - \frac{1}{\varepsilon_X \rho_X})\Big\}. \tag{4}$$

A proposer offers a positive amount if $\varepsilon_X \rho_X > 1$, which means that he or she is reasonably concerned about reciprocity in the unintentional case. No matter how large $\varepsilon_X \rho_X$ is, the offer is never greater than half. Equation (4) corresponds to the second term in (2), i. e. the possible 'voluntary offer' in the ultimatum game, where $\varepsilon_X \rho_X$ replaces ρ_X. Here, the receiver has no choice but to accept and therefore the outcome is 'unintentional' in the sense of Falk and Fischbacher (2001); this is reflected by a 'discounting' of reciprocity parameter ρ_X by pure equity concern ε_X in the dictator game. A given agent always offers weakly *less* in the dictator game than in the ultimatum game, because $0 \leq \varepsilon_X \leq 1$.

Agent A plays the dictator game with agent B and is assigned the proposer role with probability 0.5. His or her expected material payoff $\pi^*_{A_D}$ in equilibrium is then given by:

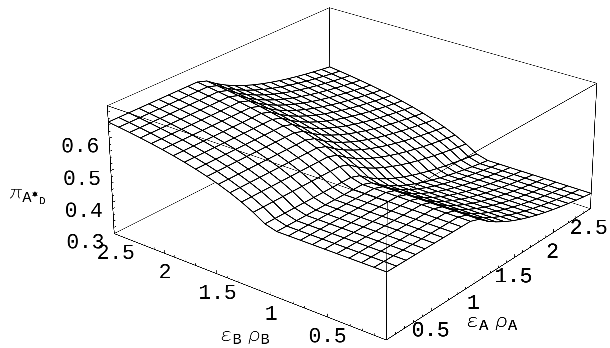

FIG. 4 Expected payoff $\pi^*_{A_D}$ of agent A with parameters ε_A and ρ_A in the dictator game when matched against an agent B with parameters ε_B and ρ_B

$$\pi^*_{A_D}(\varepsilon_A,\rho_A,\varepsilon_B,\rho_B) = \begin{cases} \frac{1}{2} & \text{if} \quad \varepsilon_A\rho_A \leq 1 \text{ and } \varepsilon_B\rho_B \leq 1 \qquad \text{(I)} \\ \frac{1}{4} + \frac{1}{4\varepsilon_A\rho_A} & \text{if} \quad \varepsilon_A\rho_A > 1 \text{ and } \varepsilon_B\rho_B \leq 1 \qquad \text{(II)} \\ \frac{3}{4} - \frac{1}{4\varepsilon_B\rho_B} & \text{if} \quad \varepsilon_A\rho_A \leq 1 \text{ and } \varepsilon_B\rho_B > 1 \qquad \text{(III)} \\ \frac{1}{2} + \frac{1}{4\varepsilon_A\rho_A} - \frac{1}{4\varepsilon_B\rho_B} & \text{if} \quad \varepsilon_A\rho_A > 1 \text{ and } \varepsilon_B\rho_B > 1. \qquad \text{(IV)} \end{cases} \tag{5}$$

We have four cases, illustrated by figure 4. In cases (II) and (III) either of the two agents offers a positive amount to the other agent when he or she is in the role of proposer. The expected payoff for such ‘generous’ behavior is smaller than in case (I), and strictly decreasing in the agent’s parameters ρ and ε. Cases (II) and (III) are hence unstable – evolution will shift preferences towards case (I).

In case (IV) *both* agents as proposers share the pie with the receiver; their payoff need not be smaller than in case (I). Still, material payoff is decreasing in each agent’s individual parameters ρ and ε, so again a payoff-monotonic evolutionary process drives agents’ preferences towards case (I).

For parameter constellations in case (I) both agents are only weakly interested in fairness; both offer $c^* = 0$ as proposers. The case applies as long as the product of the parameters for reciprocity ρ and the pure equity-concern ε is smaller or equal to unity for both agents. All such parameter constellations are behaviorally equivalent, so that any of them corresponds to an *evolutionary stable state.*[22] Constellations falling into case (I) are the only stable ones.

This implies that in the long run agents offer zero in the dictator game and

[22]Here, we refer to stability in the spirit of neutrally stable strategies and Lyapunov dynamic stability (see e. g. Weibull 1995).

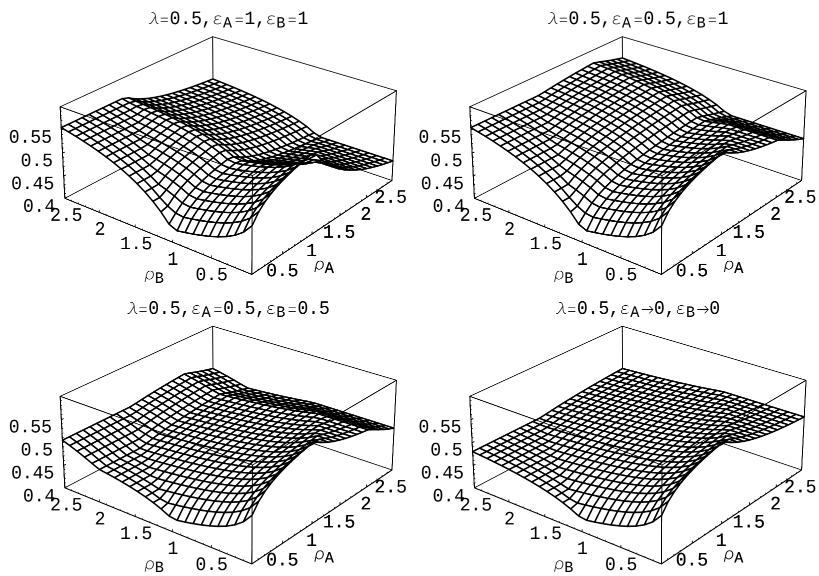

FIG. 5 Expected payoff π_A^* of agent A with parameters ε_A and ρ_A in the 'game of life' when matched against an agent B with parameters ε_B and ρ_B for $\lambda = 0.5$

thus cannot be distinguished from selfish payoff maximizers. Still, the model allows agents to have a preference for fairness up to a level that does not affect behavior in the dictator game. The 'upper bound' on this hidden fairness is given by $\rho \cdot \varepsilon \leq 1$; any combination of concern for reciprocity ρ and pure concern for equity ε that satisfies this condition can persist in the population if only the dictator game is played.

5.3. The Mixed Environment

Let us now consider the stylized 'game of life' in which agents get randomly involved either in a ultimatum or a dictator game. The payoff π_A^* of agent A facing agent B given the probability $\lambda \in [0,1]$ to play the ultimatum game and probability $1-\lambda$ to play the dictator game is

$$\pi_A^*(\varepsilon_A, \rho_A, \varepsilon_B, \rho_B, \lambda) = \lambda \cdot \pi_{A_U}^*(\rho_A, \rho_B) + (1-\lambda) \cdot \pi_{A_D}^*(\varepsilon_A, \rho_A, \varepsilon_B, \rho_B). \qquad (6)$$

This expected material payoff function is illustrated for $\lambda = 1/2$ and several values of ε_A and ε_B in figure 5.[23] Total expected payoff is increasing in payoff $\pi_{A_U}^*(\cdot)$ from the ultimatum game and the payoff $\pi_{A_D}^*(\cdot)$ from the dictator game. Hence

[23]Note the saddle points (1,1) and (2,2) in the cases of $\varepsilon_A = \varepsilon_B = 1$ and $\varepsilon_A = \varepsilon_B = 0.5$, respectively, and that a saddle point – corresponding to an evolutionary stable state if ε were fixed (or is restricted as in section 5.4) – does not exist for $\varepsilon_A = \varepsilon_B \approx 0$.

evolution would favor preferences that maximize $\pi^*_{A_U}(\cdot)$ and $\pi^*_{A_D}(\cdot)$ at the same time if this were possible.

Indeed, this is possible for the Falk-Fischbacher preferences considered here: The maximum payoff in the dictator game is reached for *any* parameter constellation with $\rho \cdot \varepsilon \leq 1$. The pure equity concern parameter ε can always totally compensate a high reciprocity concern ρ by being smaller or equal to $1/\rho$. So, the 'optimal', material payoff-maximizing preferences in the dictator game can be reached for any level of ρ. In the light of our earlier analysis the evolutionary limit outcome in the mixed environment is hence a diverging reciprocity parameter ρ and convergence of the pure equity concern parameter ε to zero such that $\rho \cdot \varepsilon \leq 1$.

In the limit, agents behave very fairly whenever their counterpart can reciprocate, but show no pure concern for equity at all. The long-run results are equal splits in the ultimatum game and full appropriation of the pie in the dictator game, i. e. we obtain a simple superimposition of the long-run outcomes in the separate components of our stylized 'game of life'. With parameter ρ diverging and ε converging towards zero, agents reciprocate strongly in situations that imply intentionality but do not care at all about fairness when it comes to situations where intentions play no role.

5.4. Restricted Parameters in the Mixed Environment

With two free parameters and two games, it is not surprising – though neither trivial – that evolution brings about 'optimal' behavior for isolated dictator and ultimatum game environments also in the combined habitat. While our agents are unable to have different sets of parameters for reciprocity and pure equity concern in ultimatum and dictator game, respectively, i. e. they cannot discriminate directly between the games, they succeed to *indirectly* discriminate based on distinctive features of the games (here: intentionality). It can be questioned whether the degrees of freedom in the social component of our preferences in reality match or exceed the different classes of social interaction.[24] Humans seem to adjust behavior to a specific situation only to some extent. It therefore seems interesting and reasonable to limit nature's freedom in shaping agents' preferences[25] by restricting the possible range of the pure outcome-concern parameter ε, which applies in unintentional situations.

One practical possibility is to impose an *exogenous lower bound* $\varepsilon_l > 0$ for parameter ε, so that $\varepsilon_l \leq \varepsilon \leq 1$ is required instead of $0 \leq \varepsilon \leq 1$. This situation

[24]Exogenous costs of discrimination by a given agent's (global) preferences would probably play a central role in a theoretical investigation of this relationship.

[25]The available preferences are already limited by the specific family of utility functions chosen by us. See Dekel, Ely, and Yilankaya (1998) for a model with finite strategy space and no such restriction.

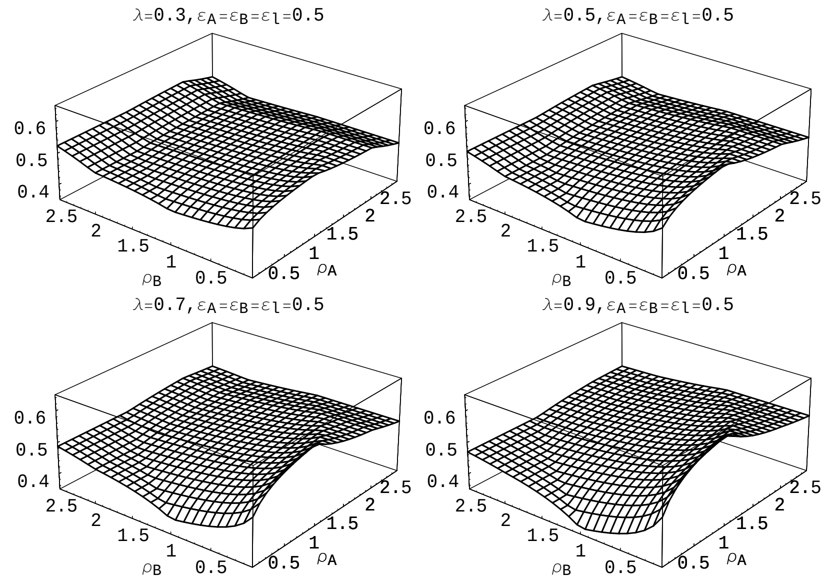

FIG. 6 Expected payoff π_A^* of agent A with parameters $\varepsilon_A = \varepsilon_l = 0.5$ and ρ_A in the 'game of life' when matched against an agent B with parameters $\varepsilon_B = \varepsilon_l = 0.5$ and ρ_B for several values of λ

differs from the above in that a strong concern for reciprocity cannot completely be blanked out in the dictator game by a low ε. An agent's behavior in dictator-realizations of the 'game of life' is thus no longer *de facto* independent from that in ultimatum-realizations.

Given the bound ε_l, the dictator game continues to induce persistent downward pressure on parameter ε for $\rho \geq 1/\varepsilon_l$. Therefore, for $\rho \geq 1/\varepsilon_l$ the evolutionary stable outcome involves $\varepsilon = \varepsilon_l$. At the same time the ultimatum game puts persistent upward pressure on reciprocity parameter ρ. This has the following implications for ρ:

For small values of λ, the dictator game is 'important' enough in agents' lives to imply $\rho \cdot \varepsilon_l = 1$ or $\rho = 1/\varepsilon_l$. This results in a zero offer in the dictator game and

$$c = \frac{1 + 3\frac{1}{\varepsilon_l} - \sqrt{1 + 6\frac{1}{\varepsilon_l} + \frac{1}{\varepsilon_l^2}}}{4\frac{1}{\varepsilon_l}} \tag{7}$$

in the ultimatum game. For example, a value of $\varepsilon_l = 0.25$ implies a split of $(\pi_X, \pi_Y) \approx (0.59, 0.41)$ or $\varepsilon_l = 0.125$ implies $(\pi_X, \pi_Y) \approx (0.55, 0.45)$ in the ultimatum game in the long run. The imposed restriction thus slightly reduces equitableness in ultimatum offers without consequences for selfish behavior in the dictator game.

The influence of parameter λ on the payoffs is illustrated by figure 6. The saddle

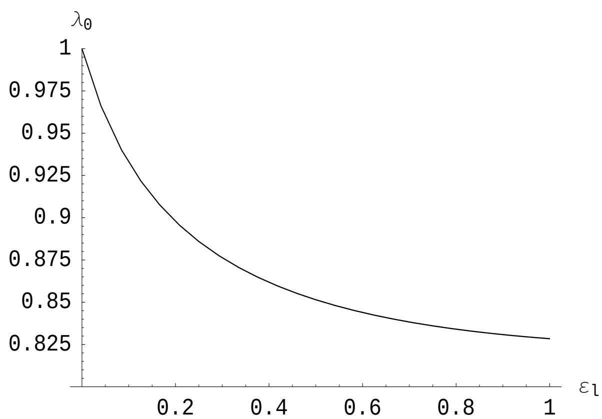

FIG. 7 The critical value $\lambda_0(\varepsilon_l)$ at which $\rho = 1/\varepsilon_l$ ceases to be an evolutionary stable point

point at $(\rho_A, \rho_B) = (2, 2)$, corresponding to a stable state with $\rho = 1/\varepsilon_l$, is barely visible for $\lambda = 0.7$ and no longer exists for $\lambda = 0.9$. In fact, there is a *critical level* of λ above which $(\rho_A, \rho_B) = (1/\varepsilon_l, 1/\varepsilon_l)$ is no longer stable and ρ diverges. This critical level can be calculated by analyzing expected payoff for values of $\rho > 1/\varepsilon_l$; if it is lower (higher) than for $\rho = 1/\varepsilon_l$ the latter combination is (is not) stable. Constellations $\rho > 1/\varepsilon_l$ belong to case (c) in the ultimatum game and case (IV) – involving positive offers by both agents – in the dictator game. The material payoff of agent A in the 'game of life' is therefore

$$\begin{aligned}\pi_A^*(\varepsilon_A, \rho_A, \varepsilon_B, \rho_B, \lambda) &= \lambda\left[\frac{3\rho_A + 1 - \sqrt{1 + 6\rho_A + \rho_A^2}}{8\rho_A} + \frac{\rho_B - 1 + \sqrt{1 + 6\rho_B + \rho_B^2}}{8\rho_B}\right] \\ &\quad +(1-\lambda)\left[\frac{1}{2} + \frac{1}{4\varepsilon_A\rho_A} - \frac{1}{4\varepsilon_B\rho_B}\right]. \qquad (8)\end{aligned}$$

The marginal payoff to agent A from greater ρ_A is

$$\frac{\partial\pi_A^*(\varepsilon_A, \rho_A, \varepsilon_B, \rho_B, \lambda)}{\partial\rho_A} = \frac{\varepsilon_A\lambda(1 + 3\rho_A) + (2\lambda - 2 - \rho_A\lambda)\sqrt{1 + 6\rho_A + \rho_A^2}}{8\varepsilon_A\rho_A\sqrt{1 + 6\rho_A + \rho_A^2}} \qquad (9)$$

and can be analyzed for $\rho_A > 1/\varepsilon_l$ and $\varepsilon_A = \varepsilon_l$ in order to find out at which critical level λ_0 the combination of selfish dictator offers and (moderately) fair ultimatum offers becomes unstable. A's marginal payoff is negative for values of λ up to the critical level

$$\lambda_0(\varepsilon_l) = \frac{2\sqrt{1 + \frac{6}{\varepsilon_l} + \frac{1}{\varepsilon_l^2}}}{3 + \varepsilon_l + (2 - \varepsilon_l)\sqrt{1 + \frac{6}{\varepsilon_l} + \frac{1}{\varepsilon_l^2}}} \qquad (10)$$

and then positive. For $\lambda \geq \lambda_0(\varepsilon_l)$ evolution results in diverging parameter ρ because in this case and for $\rho_A > 1/\varepsilon_l$ and $\rho_B > 1/\varepsilon_l$ the expected payoff $\pi_A^*(\cdot)$ is strictly increasing in ρ_A. As illustrated in figure 7, $\lambda_0(\varepsilon_l)$ is strictly decreasing in the exogenous lower bound on pure equity concern ε_l.[26]

So when the ultimatum game is played *very frequently* compared to the dictator game ($\lambda \geq \lambda_0$), the reciprocity parameter ρ diverges and the pure outcome concern resides at $\varepsilon = \varepsilon_l$. This implies equal splits in both games in the long run.[27] Loosely speaking, if agents interact in comparatively many 'private' situations captured by the ultimatum game compared to few 'market' situations reflected by the dictator game, they develop a strong notion of fairness.

6. COMPARISON WITH GÜTH AND NAPEL (2002)

Recall that the crucial difference between this model and that of Güth and Napel (2002) is the considered class of preferences. Agents' sense of fairness is limited to inequity aversion in the paper of Güth and Napel, while here a partly intention-based preference for reciprocation is investigated.

Results of the separate analysis of ultimatum and dictator game are qualitatively the same. In the *ultimatum game* agents in the role of the responder benefit from inequity aversion as well as from reciprocation, because the proposer anticipates that a small offer might be rejected. Evolution leads agents who exclusively play the ultimatum game to such high levels of inequity aversion or reciprocation, that in the long run a rather equitable or even equal split is reached (this depends on the precise utility specification).

In the *dictator game*, in contrast, both inequity aversion and reciprocal behavior are detrimental to an agent's average success, as it leads him or her to voluntarily give part of the pie away. Still, in Güth and Napel's investigation as well as this one, agents can be inequity averse to any degree that fails to actually affect proposer behavior. In the analysis by Güth and Napel, this upper bound to stable inequity aversion of dictators is driving results for the mixed environment. Either the long-run level of inequity aversion in the ultimatum game is below this bound, resulting in no behaviorally relevant interaction between the two games, or it is above. In the latter case, the given composition of the stylized 'game of life' determines a stable level that, loosely speaking, balances marginal benefits and costs of inequity aversion. In our analysis the situation is more complicated due to the presence of two free parameters instead of one. The original model of Falk and Fischbacher (2001) allows for both: Strong reciprocation and no inequity aversion in unintentional cases at all. If the possible discrepancy between these two aspects of fairness is not restricted exogenously, the reciprocity parameters goes to infinity and agents

[26] It falls from $\lambda_0 = 1$ for $\varepsilon_l \longrightarrow 0$ to a minimum of $\lambda_0 = 2\sqrt{2} - 2 \approx 0.838$ for $\varepsilon_l = 1$.

[27] A diverging parameter ρ implies equal split in the ultimatum and dictator game as long as parameter ε does not converge towards zero. For an explanation see section 5.1 on the ultimatum game and section 5.2 on the dictator game.

behave in both games just as if they were independent from each other. This corresponds to Güth and Napel's case in which agents can explicitly condition the social component of their utility function – and hence fair behavior – on the game at hand. Here, such a moral discrimination between games is achieved *indirectly*. When 'discounting' of fairness in the *unintentional* case is limited, the outcome resembles qualitatively the one of Güth and Napel's main case of game-independent equity aversion.

It is a common result of either analysis that a high share of ultimatum games affects evolutionary stable preferences. Güth and Napel find moderate increases in the stable level of inequity aversion after raising λ, while here the reciprocity parameter diverges once a critical point is reached. Coincidentally, the minimal level at which the amount of ultimatum games played affects the outcome is similar in both studies, about 80% in Güth and Napel (2002) and around 85% for a wide range of constellations here; below these thresholds, the contrasting long-run outcomes for dictator and ultimatum games are independent of the precise composition of the stylized 'game of life'. In summary, both specifications of the utility function evolution lead to qualitatively similar outcomes in ultimatum, dictator and mixed environments. The impact that an explicit inclusion of reciprocal behavior has on the evolutionary stable outcome is small in the considered class of games if the flexibility increase from introducing an additional parameter is limited.

7. CONCLUDING REMARKS

In general, psychological fairness models explain experimental results better than more tractable equity-based models, which in turn perform better than the default assumption of monetary payoff maximization. This success comes at the expense of additional free parameters, which have do be defended against accusations of ad hoc 'game fitting', and often more complicated analysis with sometimes ambiguous results. Since parameters *can* be such that an agent, in fact, has traditional materialistic preferences, indirect evolutionary analysis provides a useful theoretical benchmark. If under plausible modelling assumptions evolutionary forces selected parameters that imply non-degenerate fairness preferences, this would corroborate the post-experimental econometric estimations.

The relevance of fairness preference a priori can be expected to vary with the decision situations at hand even if these are restricted to simple distribution tasks. Our analysis suggests that preference for fair behavior has sound evolutionary reasons, but – in line with experimental observations – more likely plays a significant role in non-degenerate games with punishment opportunities, such as the ultimatum game, than in what is basically a one-player decision problem like the dictator game. If Nature permits players to condition the social component of their preferences on different games (as investigated by Güth and Napel 2002) or at least indirectly on differences in games, the same agents can exhibit a pronounced sense of fairness in one type of social interaction while they are entirely selfish in another one. This is true even though preferences are assumed to evolve simultaneously in a multi-game environment.

If Nature imposes physical or psychological restrictions on the variance in agents' social attitudes across different games or different dimensions of fairness, the precise composition of the stylized 'game of life' faced by agents has an impact. In our model, evolutionary pressure towards more reciprocal preferences created by the ultimatum game is strong enough to induce quite equitable behavior (offers above 40%) in this game independently of how frequently it is played. In contrast, the success of evolutionary pressure towards more selfish preferences created by the dictator game is sensitive to the share of dictatorial decision problems: If agents most of the time face the more fairness-encouraging ultimatum game, they will eventually feel it in their interest to be generous even as dictators. Whether this asymmetry – a great share of fairness-encouraging games eliminates selfish behavior in materialism-encouraging games, while the reverse is not true – in response to possible changes in the composition of the environment is a more general phenomenon or an artifact of our (and also Güth and Napel's) model specification remains to be seen.

We have studied environmental and psychological determinants of material-payoff maximization vs. social preferences in a particularly simple two-game environment. In our view, this is an improvement compared to the usual analysis of preferences in a single game. It highlights the importance of possibly unconscious links between behavioral modes in different classes of interaction. Whether human agents indeed face a binding restriction in their ability to discriminate between the fairness implications of similar play in ultimatum and dictator games, reflected by the reciprocity and pure equity concern parameters ρ and ε in our model, is an empirical question. Experimental evidence on the ultimatum game is consistent with both the unrestricted (section 5.3) and restricted evolution of ρ and ε (section 5.4); evidence on the dictator game supports the restricted view, though not our prediction of equal splits.[28]

It seems desirable to us to extend our 'game of life' to more than only ultimatum and dictator games. First, an environment with more than two component games would create a natural endogenous restriction on the two parameters of the considered Falk-Fischbacher reciprocal preferences – preventing the full 'specialization' observed in section 5.3 without exogenous restrictions as in section 5.4. Second, reciprocal preferences are proposed as explanations for empirical observations of

[28]Dictator offers in Güth and Napel (2002) continuously rise from zero to moderate positive levels (assuming that marginal disutility of inequality is increasing).

a considerable range of games, including e. g. trust and gift-exchange games, public good games, and variants of the ultimatum game such as the best-shot game. Analysis of the stability of reciprocal preferences should include these.

Another very desirable extension would be to relax our a priori restriction of the class of feasible preferences to those of the Falk-Fischbacher type.[29] These include the neoclassical reference point of selfish materialism as well as salient other special cases (e. g. purely intention-based fairness or a general equity concern). They admittedly exclude many other types of agents, which are competing with the traditional homo oeconomicus for economists' attention, too.

REFERENCES

Andreoni, J. and J. H. Miller (2002). Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica 70*, 737–753.

Benaim, M. and J. W. Weibull (2003). Deterministic approximation of stochastic evolution in games. *Econometrica 71*(3), 873–903.

Berg, J., J. Dickhaut, and K. McCabe (1995). Trust, reciprocity and social history. *Games and Economic Behavior 10*, 122–142.

Berninghaus, S. K., W. Güth, and H. Kliemt (2003). From teleology to evolution – Bridging the gap between rationality and adaptation in social explanation. *Journal of Evolutionary Economics 13*(4), 385–410.

Bolton, G. and A. Ockenfels (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review 90 (1)*, 166–193.

Camerer, C. F. and R. H. Thaler (1995). Ultimatums, dictators, and manners. *Journal of Economic Perspectives 9*, 209–219.

Cameron, L. A. (1999). Raising the stakes in the ultimatum game: Experimental evidence from Indonesia. *Economic Inquiry 37*(1), 47–59.

Charness, G. and M. Rabin (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics 117*, 817–869.

Daws, R. M. and R. H. Thaler (1988). Cooperation. *Journal of Economic Perspectives 2*, 187–197.

Dekel, E., J. Ely, and O. Yilankaya (1998). Evolution of preferences. mimeo.

[29]Giving up *any* restriction in the spirit of Dekel, Ely, and Yilankaya (1998) would in our view go to far: It implicitly allows for indirect moral discrimination between the components of any mixed habitat. One thus returns to the unrealistic superimposition of possibly many different preference types that are each tailor-made for a specific single-game environments.

Dufwenberg, M. and G. Kirchsteiger (2001). A theory of sequential reciprocity. Discussion Paper 37, Version July 2001, Tilburg University, CentER for Economic Research.

Falk, A., E. Fehr, and U. Fischbacher (2000). Informal sactions. Working Paper 59, Institute for Empirical Research in Economics, University of Zürich.

Falk, A., E. Fehr, and U. Fischbacher (2003). On the nature of fair behavior. *Economic Inquiry 41*(1), 20–26.

Falk, A. and U. Fischbacher (2001). A theory of reciprocity. Working Paper 457, CESifo, München.

Falk, A. and S. Gächter (2002). Reputation and reciprocity – Consequences for the labour relation. *Scandinavian Journal of Economics 104*, 1–26.

Fehr, E. and S. Gächter (1998). Reciprocity and economics. the economic implications of homo reciprocans. *European Economic Review 42*, 845–859.

Fehr, E. and S. Gächter (2000). Cooperation and punishment in public goods experiments. *American Economic Review 90 (4)*, 980–994.

Fehr, E., S. Gächter, and G. Kirchsteiger (1996). Reciprocal fairness and noncompensating wage differentials. *Journal of Institutional and Theoretical Economics 152(4)*, 608–640.

Fehr, E. and K. Schmidt (1999). A theory of fairness, competition and cooperation. *Quarterly Journal of Economics August*, 817–868.

Fehr, E. and K. M. Schmidt (2002). Theories of fairness and reciprocity – Evidence and economic application. In M. Dewatripont, L. P. Hansen, and S. J. Turnovsky (Eds.), *Advances in Economic Theory, Eigth World Congress of the Econometric Society*. Cambridge, UK: Cambridge University Press. Forthcoming.

Forsythe, R., J. Horowitz, N. Savin, and M. Sefton (1994). Fairness in simple bargaining games. *Games and Economic Behavior 6*, 347–369.

Geanakoplos, J., D. Pearce, and E. Stacchetti (1989). Psychological games and sequential rationality. *Games and Economic Behavior 1*, 60–79.

Güth, W., H. Kliemt, and S. Napel (2003). Wie du mir, so ich dir! – Evolutionäre Modellierungen. In M. Held, G. Kubon-Gilke, and R. Sturn (Eds.), *Jahrbuch Normative und Institutionelle Grundfragen der Ökonomik, Band 2: Experimentelle Ökonomik*, pp. 113–139. Marburg: Metropolis-Verlag.

Güth, W. and S. Napel (2002). Inequality aversion in a variety of games – An indirect evolutionary analysis. Discussion Paper on Strategic Interaction 2002-23, Max Planck Institute for Research into Economic Systems, Strategic Interaction Group.

Güth, W., R. Schmittberger, and B. Schwarze (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization 3*, 367–388.

Güth, W. and R. Tietz (1990). Ultimatum bargaining behavior – A survey and comparison of experimental results. *Journal of Economic Psychology 11*, 417–449.

Harrison, G. W. and J. Hirshleifer (1989). An experimental evaluation of weakest link/best shot models of public goods. *Journal of Political Economy 97*, 201–225.

Huck, S. and J. Oechssler (1999). The indirect evolutionary approach to explaining fair allocations. *Games and Economic Behavior 28*, 13–24.

Klein, A. (2000). *Reciprocity, Endogenous Incomplete Contracts, and the Role of Options in Law*. Berlin: dissertation.de.

Koçkesen, L., E. Ok, and R. Sethi (2000a). Evolution of interdependent preferences in aggregative games. *Games and Economic Behavior 31*, 303–310.

Koçkesen, L., E. Ok, and R. Sethi (2000b). The strategic advantage of negatively interdependent preferences. *Journal of Economic Theory 92(2)*, 274–299.

Ledyard, J. (1995). Public goods: A survey of experimental research. In J. Kagel and A. E. Roth (Eds.), *Handbook of Experimental Economics*. Princeton, New Jersey: Princeton University Press.

Oechssler, J. and F. Riedel (2002). On the dynamic foundation of evolutionary stability in continuous models. *Journal of Economic Theory 107*, 223–252.

Poulsen, A. and O. Poulsen (2003). Evolution of reciprocal, materialistic and altruistic preferences in an environment of prisoner's dilemma games. mimeo, Department of Economics, Aarhus School of Business.

Prasnikar, V. and A. E. Roth (1992). Considerations of fairness and strategy: Experimental data from sequential games. *Quarterly Journal of Economics 107(3)*, 865–888.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review 83 (5)*, 1281–1302.

Roth, A. E. (1995). Bargaining experiments. In J. Kagel and A. E. Roth (Eds.), *Handbook of Experimental Economics*. Princeton, New Jersey: Princeton University Press.

Samuelson, L. (2001). Introduction to the evolution of preferences. *Journal of Economic Theory 97*, 225–230.

Sethi, R. and E. Somanathan (2001). Preference evolution and reciprocity. *Journal of Economic Theory 97*, 273–297.

Slonim, R. and A. E. Roth (1998). Learning in high stakes ultimatum games: An experiment in the Slovak Republic. *Econometrica 66*(3), 569–596.

Stahl, D. O. and E. Haruvy (2002). Level-n bounded rationality on a level playing field of sequential games. mimeo, University of Texas at Austin and Dallas.

Thaler, R. H. (1988). The ultimatum game. *Journal of Economic Perspectives II, No. 4*, 195–206.

Weibull, J. W. (1995). *Evolutionary Game Theory.* Cambridge, Massachusetts: The MIT Press.

Nr.	Author	Title
03-28	Alexander Zimper	Uniqueness Conditions for Point-Rationalizable Solutions of Games with Metrizable Strategy Sets
03-27	Jürgen Eichberger David Kelsey	Sequential Two-Player Games with Ambiguity
03-26	Alain Chateauneuf Jürgen Eichberger Simon Grant	A Simple Axiomatization and Constructive Representation Proof for Choquet Expected Utility
03-25	Volker Stocké	Informationsverfügbarkeit und Response-Effects: Die Prognose von Einflüssen unterschiedlich kategorisierter Antwortskalen durch Antwortsicherheiten und Antwortlatenzen
03-24	Volker Stocké	Entstehungsbedingungen von Antwortverzerrungen durch soziale Erwünschtheit. Ein Vergleich der Prognosen der Rational-Choice Theorie und des Modells der Frame-Selektion
03-23	Daniel Schunk	Modeling the Use of Nonrenewable Resources Using a Genetic Algorithm
03-22	Brian Deal Daniel Schunk	Spatial Dynamic Modeling and Urban Land Use Transformation: An Ecological Simulation Approach to Assessing the Costs of Urban Sprawl
03-21	Thomas Gschwend Franz Urban Pappi	Stimmensplitting und Koalitionswahl
03-20	Thomas Langer Martin Weber	Does Binding or Feeback Influence Myopic Loss Aversion - An Experimental Analysis
03-19	Peter Albrecht Carsten Weber	Asset/Liability Management of German Life Insurance Companies: A Value-at-Risk Approach in the Presence of Interest Rate Guarantees
03-18	Markus Glaser	Online Broker Investors: Demographic Information, Investment Strategy, Portfolio Positions, and Trading Activity
03-17	Markus Glaser Martin Weber	September 11 and Stock Return Expectations of Individual Investors

Nr.	Author	Title
03-16	Siegfried K. Berninghaus Bodo Vogt	Network Formation and Coordination Games
03-15	Johannes Keller Herbert Bless	When negative expectancies turn into negative performance: The role of ease of retrieval.
03-14	Markus Glaser Markus Nöth Martin Weber	Behavioral Finance
03-13	Hendrik Hakenes	Banks as Delegated Risk Managers
03-12	Elena Carletti	The Structure of Bank Relationships, Endogenous Monitoring and Loan Rates
03-11	Isabel Schnabel	The Great Banks‘ Depression - Deposit Withdrawals in the German Crisis of 1931
03-10	Alain Chateauneuf Jürgen Eichberger Simon Grant	Choice under Uncertainty with the Best and Worst in Mind: Neo-additive Capacities.
03-09	Peter Albrecht Carsten Weber	Combined Accumulation- and Decumulation-Plans with Risk-Controlled Capital Protection
03-08	Hans-Martin von Gaudecker Carsten Weber	Surprises in a Growing Market Niche - An Evaluation of the German Private Annuities Market
03-07	Martin Hellwig Martin Weber	Overconfidence and Trading Volume
03-06	Markus Glaser Thomas Langer Martin Weber	On the trend recognition and forecasting ability of professional traders
03-05	Geschäftsstelle	Jahresbericht 2002
03-04	Oliver Kirchkamp Rosemarie Nagel	No imitation - on local and group interaction, learning and reciprocity in prisoners break
03-03	Michele Bernasconi Oliver Kirchkamp Paolo Paruolo	Expectations and perceived causality in fiscal policy: an experimental analysis using real world data

Nr.	Author	Title
03-02	Peter Albrecht	Risk Based Capital Allocation
03-01	Peter Albrecht	Risk Measures
02-51	Peter Albrecht Ivica Dus Raimond Maurer Ulla Ruckpaul	Cost Average-Effekt: Fakt oder Mythos?
02-50	Thomas Langer Niels Nauhauser	Zur Bedeutung von Cost-Average-Effekten bei Einzahlungsplänen und Portefeuilleumschichtungen
02-49	Alexander Klos Thomas Langer Martin Weber	Über kurz oder lang - Welche Rolle spielt der Anlagehorizont bei Investitionsentscheidungen?
02-48	Isabel Schnabel	The German Twin Crisis of 1931
02-47	Axel Börsch-Supan Annamaria Lusardi	Saving Viewed from a Cross-National Perspective
02-46	Isabel Schnabel Hyun Song Shin	Foreshadowing LTCM: The Crisis of 1763
02-45	Ulrich Koch	Inkrementaler Wandel und adaptive Dynamik in Regelsystemen
02-44	Alexander Klos	Die Risikoprämie am deutschen Kapitalmarkt
02-43	Markus Glaser Martin Weber	Momentum and Turnover: Evidence from the German Stock Market
02-42	Mohammed Abdellaoui Frank Voßmann Martin Weber	An Experimental Analysis of Decision Weights in Cumulative Prospect Theory under Uncertainty
02-41	Carlo Kraemer Martin Weber	To buy or not to buy: Why do people buy too much information?
02-40	Nikolaus Beck	Kumulation und Verweildauerabhängigkeit von Regeländerungen