Inequality Aversion in a Variety of Games – An Indirect Evolutionary Analysis[*]

Werner Güth
Max Planck Institute
for Research into Economic Systems
Kahlaische Straße 10
D-07745 Jena
Germany
gueth@mpiew-jena.mpg.de

Stefan Napel
University of Hamburg
Dept. of Economics
Von-Melle-Park 5
D-20146 Hamburg
Germany
napel@econ.uni-hamburg.de

2$^{\text{nd}}$ revision, June 2005

The indirect evolutionary approach integrates forward-looking evaluation of opportunities and adaptation in the light of the past. Subjective motivation determines behavior, but long-run evolutionary success of motivational types depends on objective factors only. This can justify intrinsic aversion to inequality in reward allocation games. Whereas earlier analysis was restricted to specific games, this paper considers a more complex environment comprising different games which – studied in isolation – yield opposite implications for the survival of inequality aversion. Persistent divergence between intrinsic motivation and true material success is possible depending on the definition of inequality aversion as well as on agents' ability to discriminate between games.

1. INTRODUCTION

In traditional microeconomic analysis decision alternatives are selected by economic men based on their anticipated consequences, using preferences which are fixed and given. Other social analysis and evolutionary biology, in contrast, focus on the shadow of the past. Propensities to act in this or that way are explained by the social or biological environment and the past success of possible strategies (mutants) in it. Studying the evolution of preferences offers the chance to combine forward-looking deliberation (the shadow of the future) and path dependence (the shadow of the past). What evolves is not behavior itself, as typically assumed in evolutionary biology and (direct) evolutionary game theory, but its determinants.[1]

[*] We gratefully acknowledge the constructive comments made by A. Poulsen and two anonymous referees.

Preferences of economic agents have many degrees of freedom in neoclassical decision theory. They need not – though this is often regarded as economic behavior per se – be egoistic and purely materialistic. This option has frequently been exercised in the context of surplus division (see Bolton 1991, Rabin 1993, Kirchsteiger 1994, Fehr and Schmidt 1999, Bolton and Ockenfels 1999 and 2000 to mention just a few), in particular to explain rejection of unfair offers and equitable proposals in ultimatum experiments.

Rationalization of experimental observations is generally possible ex post by positing suitable preferences. However, fitting utility functions with empirical results is rather ad-hoc and, at least as far as the common knowledge of such idiosyncratic preferences is concerned, quite questionable. To be more than merely neoclassical repairs of a priori assumptions about behavior they should be robust in an empirical and a theoretical sense. First, they should explain more than one particular set of observations (see the systematic attempts by Bolton and Ockenfels 2000 and Fehr and Schmidt 1999). Second, they should not be in contradiction with the physical necessity (and observable tendency) to strive and compete for material rewards in a world of scarce resources. If agents exhibit non-opportunistic preferences, this, if sustainable, should not imply a significant and persistent disadvantage.

Evolution of agents' utility functions has so far been studied in a highly artificial world;[2] a very stylized single-game environment is taken to determine biological or social success.[3] A rare strain of virus may take over in a monoculture, though it would never have a chance in a natural ecosystem (like otherwise non-existent species, e. g. non-flying birds, in isolated island habitats). With this in mind, present applications of evolutionary game theory that study (direct or indirect) evolution for just one specific game are still inconclusive.

Human behavior (but not only – see, for instance, de Waal 1998) or, rather, its basic determinants are not game-specific. Men act similarly in entire, quite general classes of games. People usually do not mind to opportunistically exploit others in market interaction. But they are reluctant to do so when an experimental setting suggests a private affair (even though it is anonymous and single-shot). To explain how such general indications can evolve one has to study evolution of behavior or behavioral determinants not for just one game but for the 'game of life', encompassing different game types between which agents may or may not be able to discriminate.[4]

[1] For related studies see the contributions collected in the symposium in *Journal of Economic Theory*, Vol. 97, Nr. 2, (2001).

[2] See e. g. Huck and Oechssler (1999), Koçkesen, Ok, and Sethi (2000a, 2000b), and Sethi and Somanathan (2001).

[3] An exception is Poulsen and Poulsen (2005).

[4] We associate with 'game of life' the general research program of studying evolution for ever

We do not know how the ‘game of life’ that is shaping our preferences can adequately be modelled. More modestly, we want to illustrate

- how (indirect) evolution can be analyzed for a variety of game types rather than for one specific game and
- that conclusions about stable preferences depend in an instructive way on the compound strategic environment.

For the sake of specificity we focus on one possible determinant of game playing behavior,[5] namely other-regarding preferences in the form of *inequality aversion* (see Bolton 1991, Fehr and Schmidt 1999, Bolton and Ockenfels 1999 and 2000, and Possajennikov 2000). Our interest is primarily to extend analysis into another direction, namely to environments composed of several games, and not to propagate inequality aversion.

We investigate whether inequality aversion evolves for a mixed environment which exhibits at least some of the variety characterizing real social environments. Our analysis focuses on distribution conflicts among two parties and two well-known procedures for solving them: dictatorial reward allocation and ultimatum bargaining. These two games have different ramifications concerning the evolutionary (dis-)advantage of inequality aversion, and thus allow the analysis of very general trade-offs concerning social preferences and the crucial issue of game specificity in a fairly simple setting. Moreover, Bolton and Ockenfels (2000) demonstrate with their ERC model that “many facets of behavior, over a wide class of games, can be deduced from [these] two . . . most elementary games” (p. 188). Namely, dictator offer and lowest accepted offer in ultimatum bargaining define flash points that predict behavior in other bargaining, market, and social dilemma games surprisingly well. Thus our quite specific study may have rather general implications.

Section 2 introduces our model, the compound material environment in which evolution operates and the different possibilities of its subjective evaluation by agents. Section 3 studies evolution of inequality aversion separately for each game

richer and more realistic habitats and of going beyond analysis of only game-specific evolution. The term has been used in diverse contexts in the literature, ranging from J. H. Conway’s cellular automata (Gardner 1970) to predator-prey systems and repeated Prisoner’s Dilemma games (see e. g. Sigmund 1995). Binmore (1994, 1998) has made it the central concept of his bargaining-theoretic analysis of social contracts.

[5] This avoids considerably more complex analysis which would be necessary if, for example, game-specific ‘commitment preferences’ (see Ok and Vega-Redondo 2001, Ex. 1) were also considered. The quite general instability of individualistic preferences under perfect observability is nicely illustrated by preferences which ‘commit’ an agent to playing his or her best strategy (anticipating that his or her opponent plays a best response to it). To have such tailor-made preferences for all possible games, however, seems unrealistic. – For the analysis of an unrestricted, albeit finite preference domain see Dekel, Ely, and Yilankaya (1998).

type. Then, section 4 considers the mixed environment. We focus on the implications of perfectly observable preferences, but also discuss alternatives. Section 5 comments on particular extensions and modifications. Notably, we look at a 3-player version of the Ultimatum Game which requires agents to trade off own material payoff as well as advantageous and disadvantageous inequality. Section 6 concludes.

2. THE MODEL

Consider a population of agents who in pairs of two can create an interpersonally comparable surplus (the 'pie'). Over and over again, pairs of agents are matched at random with a corresponding opportunity to produce and share. Let the population size be sufficiently large such that one can ignore repeated game effects, i. e. each interaction between two agents is single-shot. Agents are fully rational and act strategically given their individual preferences. The latter can (but need not) depend on other factors than own surplus share. In particular, agents may care about the *distribution* of material payoff. Preferences are considered common knowledge.

Players' equilibrium behavior, based on their subjective evaluation of feasible actions, determines material rewards. These are the unique determinant of players' reproductive (or imitative) success. Subjective preferences that yield greater than average material payoff will increase their population share, while those below average will become less frequent. As nicely phrased by Samuelson (2001, p. 226f), "Nature can thus mislead her agents, in that preferences and fitnesses can diverge, but cannot mislead herself, in that high fitness wins the day."

2.1. The Material World

Agents' environment comprises two games. The first is the *Ultimatum Game*. The agent in role X, also referred to as the *distributor*, offers a share $y \in S' \equiv [0, 1]$ of a surplus normalized to unity to the agent in role Y. The latter, also referred to as the *receiver*, either accepts or rejects this offer, in both cases ending the game. Acceptance and rejection imply the material payoff vectors $(\pi_X, \pi_Y) = (1 - y, y)$ and $(0, 0)$, respectively.

The second game that a pair of agents may play is the *Dictator Game*. It differs from the Ultimatum Game in that the receiver is only a dummy, i. e. an offer y by player X immediately results in payoffs $(\pi_X, \pi_Y) = (1 - y, y)$.

Agents are assigned to roles X and Y at random and with equal probability 0.5. Which game two agents will play in their encounter is also random with probability $\lambda \in (0, 1)$ for the Ultimatum Game and probability $1 - \lambda$ for the Dictator Game.

The material rewards π_X and π_Y measure reproductive success and are also a main object of the preferences which determine behavior. If agents respectively maximized π_X and π_Y, they would in equilibrium offer $y = 0$ as X in both Ultimatum and Dictator Games, and accept any offer as Y in the Ultimatum Game.

2.2. Intrinsic Motivation

Agents need not be concerned with their individual material reward alone. Their preferences in the roles of X and Y, represented by utility functions u_X and u_Y, are in principle only restricted by the rationality requirements of completeness and transitivity. Among the many aspects other than individual material reward that may matter in the context of bargaining, we concentrate on aversion to inequality. It can have different intensity relative to the desirability of material payoff. Moreover, it may be specified in different ways. An agent may, for instance (see Bolton 1991), suffer a disutility of ending up less well off than the other player, but be indifferent when the other player is less well off. In contrast, an agent with two-sided inequality aversion always suffers from unequal material payoffs regardless who is disadvantaged.

If agents always prefer a larger to a smaller share of the pie, an accepted equitable offer $y = 0.5$ is preferred to $y > 0.5$ by player X, and player Y is better off accepting $y = 0.5$ than rejecting it. Therefore, offers $y > 1/2$ cannot be observed in equilibrium. Any inequity of the equilibrium outcome must be to the disadvantage of player Y. We can use this to simplify agents' utility,[6] and restrict player X's strategy space to $S \equiv [0, 1/2]$. The specific functional form of incorporating inequality aversion by Y does not matter much. For the sake of specificity, we assume that utility in role Y is given by

$$u_Y(\pi_X, \pi_Y) = \pi_Y - i_Y\sqrt{\pi_X - \pi_Y} \tag{1}$$

for an agent-specific parameter $i_Y \geq 0$.

Two-sided inequality aversion affects behavior also as a distributor. It can be modelled by assuming

$$u_X(\pi_X, \pi_Y) = \pi_X - i_X\sqrt{\pi_X - \pi_Y}. \tag{2}$$

Equation (2) takes X's marginal pain from more inequality to decrease with the prevailing level of inequality. In a sense, the distributor gets more and more used to inequality the greater it is.

An alternative specification of X's utility under two-sided inequality is

$$\tilde{u}_X(\pi_X, \pi_Y) = \pi_X - \frac{i_X}{4}(\pi_X - \pi_Y)^2. \tag{2'}$$

[6] In particular, we do not include terms that capture utility loss associated with disadvantageous inequality in the distributor role and advantageous inequality as receiver.

Equation (2') takes X's marginal pain from more inequality to increase with its level. This implies that X does not care too much when the payoff difference $\Delta = \pi_X - \pi_Y \geq 0$ is small, but gets ever more sensitive to changes in inequality when Δ is large. Though many more specifications of utility in the roles of X and Y are possible,[7] we concentrate on either (1) and (2) or (1) and (2').

2.3. Discrimination between Games

Above specifications of agents' preferences allow their intrinsic motivation to depend on the role (X or Y) that is assigned to them. Characteristics i_X and i_Y can, in principle, evolve separately from each other. We will briefly consider this possibility below, but concentrate on the case where inequality aversion is a one-dimensional general characteristic so that $i_X \equiv i_Y$ for any given agent.

Independent from this, it may or may not be possible for agents to morally discriminate between several classes of interaction and to have distinct inequality aversion for each. Despite identical equilibrium outcomes, the Ultimatum and Dictator Games represent different types of interaction. The Dictator Game is degenerate in the sense that player Y has no influence. In contrast, in the Ultimatum Game both players' payoffs truly depend on their cooperation.

We will first investigate perfect discrimination. In this case, different parameters i^U and i^D (possibly further specialized to i^U_X, i^U_Y, etc.) are applied in Ultimatum and Dictator Game, respectively. Then, the case of no discrimination is studied, corresponding to a unique characteristic parameter $i \equiv i^U \equiv i^D$ for each agent. Perfect vs. imperfect (moral) discrimination between different types of interaction echoes the distinction between complete and incomplete preference information, at a purely internal level. The latter's different implications for stability of non-individualistic preferences and behavior deviating from a Nash equilibrium in the underlying objective game is emphasized by Ok and Vega-Redondo (2001) and Ely and Yilankaya (2001).[8] One can suspect a similar difference in our context or even conjecture that the evolutionary benefits of non-individualistic preferences erode as its domain is extended to more qualitatively different games. The advantage of conditioning on the given strategic situation – closely related to commitment – is

[7] For example, one could study the effect of increasing marginal disutility in the role of Y, too. Also see Güth and Napel (2003) for an explicit investigation of one-sided inequality aversion, corresponding to $i_X \equiv 0$.

[8] Ok and Vega-Redondo rigorously study evolution when agents only observe the distribution of (two types of) preferences in a finite population, but not in the particular subgroup of players they are interacting with. If the size of each subgroup (in our case: two agents) is small relative to the total population, only individualistic preferences, which perfectly reflect objective payoffs, will survive. Related, Ely and Yilankaya show that only preferences which induce play of a Nash equilibrium in objective payoffs will survive if all possible preferences over a finite set of outcomes evolve in an infinite population.

possibly dominated by disadvantages in other situations.

3. EVOLUTION OF INEQUALITY AVERSION

Let X observe Y's aversion parameter i_Y before choosing offer y. In the spirit of Güth (1995), Sethi and Somanathan (2001), and Güth, Kliemt, and Napel (2003), this could be weakened to X being aware of i_Y for a positive fraction of interactions, the availability of sufficiently accurate signals, or X having the choice to find out about his or her counterpart's preferences at sufficiently small costs.[9]

We start with an individual analysis of the Dictator and the Ultimatum Game (or the boundary cases $\lambda = 0$ and $\lambda = 1$), and in the subsequent section consider the more complex environment involving both games under different assumptions.

3.1. The Dictator Game

Formulation (2) takes marginal pain from more inequality to decrease with the prevailing level of inequality. Given any positive aversion against inequality, player X's total utility of offering y is either increasing on the entire interval $[0, 1/2]$ or U-shaped with local maxima at boundary points $y = 0$ and $y = 0.5$. Therefore, (2) implies a 'bang-bang'-solution for the optimal dictator offer

$$y^{**}(i_X^D) = \begin{cases} 0; & i_X^D \leq 0.5 \\ 0.5; & i_X^D > 0.5, \end{cases} \tag{3}$$

where we assume that the distributor resolves the tie for $i_X^D = 0.5$ in own favor.[10]

The alternative specification of X's utility, (2'), takes X's marginal pain from more inequality to increase. For i_X^D very close to zero, even maximal inequality ($\Delta = 1$) cannot produce enough interest in reducing the agent's bad conscience, i. e. the optimal dictator offer is 0. However, for i_X^D sufficiently large, there is an interior optimum resulting in dictator offer

$$\tilde{y}^{**}(i_X^D) = \max\Big\{0, \frac{i_X^D - 1}{2i_X^D}\Big\}. \tag{4}$$

In the case of utility function u_X, the expected material payoff to an agent with inequality aversion i_X^D who plays the Dictator Game with an agent with inequality

[9] For a careful analysis of the case of private information see Güth (1995), Güth and Peleg (2001), and Ok and Vega-Redondo (2001).

[10] The assumption that a receiver accepts an ultimatum offer if he is indifferent is needed for existence of a sub-game perfect equilibrium given proposal space $S = [0, 0.5]$. Its evolutionary credentials are ambiguous: If $i_Y^D = 0$ and a zero share is offered, rejecting is better in terms of relative fitness. However, if $i_Y^D > 0$ and a positive offer is made, rejection hurts in comparison with the rest of a large population.

aversion $i_X^{D\prime}$ is

$$\Pi_i^D(i_X^D, i_X^{D\prime}) = \begin{cases} \frac{1}{2}; & i_X^D, i_X^{D\prime} \leq \frac{1}{2} \\ \frac{3}{4}; & i_X^D \leq \frac{1}{2} \wedge i_X^{D\prime} > \frac{1}{2} \\ \frac{1}{4}; & i_X^D > \frac{1}{2} \wedge i_X^{D\prime} \leq \frac{1}{2} \\ \frac{1}{2}; & i_X^D, i_X^{D\prime} > \frac{1}{2}. \end{cases} \tag{5}$$

An agent with $i_X^D > \frac{1}{2}$ makes generous offers in the role of X. Any mutant with $i_X^{D\prime\prime} \leq 1/2$, who offers only 0, obtains higher expected payoff. Hence, in the long run, only agents with inequality aversion not exceeding 1/2 will be observed. Agents may differ in positive inequality aversion that is too weak to be noticed. It is symbolic at best, dominated by the strictly monotonic preference for own payoff.

In case of utility $\tilde{u}_X$ in the role of player X, we obtain

$$\tilde{\Pi}_i^D(i_X^D, i_X^{D\prime}) = \begin{cases} \frac{1}{2}; & i_X^D, i_X^{D\prime} < 1 \\ \frac{1}{2} + \frac{i_X^{D\prime}-1}{4i_X^{D\prime}}; & i_X^D < 1 \wedge i_X^{D\prime} \geq 1 \\ \frac{i_X^D+1}{4i_X^D}; & i_X^D \geq 1 \wedge i_X^{D\prime} < 1 \\ \frac{i_X^D+1}{4i_X^D} + \frac{i_X^{D\prime}-1}{4i_X^{D\prime}}; & i_X^D, i_X^{D\prime} \geq 1. \end{cases} \tag{6}$$

Again, agents with a noticeable inequality aversion – here corresponding to $i_X^D > 1$ – fare worse, ceteris paribus, than an agent whose (possible) aversion does not translate into making positive dictator offers. So, independent of the precise specification of inequality aversion, the latter is selected and will eventually predominate. Generous dictators keep less material payoff for themselves and cannot compete with their more selfish rivals.

3.2. The Ultimatum Game

Initially assume that possible inequality aversion is a general rather than role-specific disposition, i. e. $i^U \equiv i_X^U \equiv i_Y^U$. We keep the subscripts X and Y to emphasize which player is determining equilibrium offers in a given interaction between two agents with parameters i^U and $i^{U\prime}$, respectively. There are two obvious candidates for the equilibrium offer in the Ultimatum Game. First, the 'dictator offer' $y^{**}(i_X^U)$ given by (3) may be proposed and accepted.[11] This will be the case if X prefers to offer 0.5 given $i_X^U > 0.5$. Second, if i_X^U is small compared to $i_Y^{U\prime}$, the constraint in

$$\max_{0 \leq y \leq 1/2} u_X(1-y, y) \quad \text{s. t.} \quad u_Y(1-y, y) \geq 0$$

is binding. Then the optimal ultimatum offer is

$$y^*(i_Y^U) = i_Y^U \sqrt{{i_Y^U}^2 + 1} - {i_Y^U}^2, \tag{7}$$

[11] Or, for our alternative specification of X's inequality aversion, $\tilde{y}^{**}(i_X^U)$ given by (4).

and will be accepted by player Y in equilibrium.

Moreover, there exists an interesting third possibility for utility specification u_X in (2): For intermediate levels of i_X^U yielding a U-shaped function $u_X(1-y,y)$, a distributor who would maximize utility by offering $y^{**}=0$ as a dictator may have a constrained global optimum at $y^*=0.5$. The constraint $u_Y(1-y,y)\geq 0$ does not actually bind but nevertheless induces different distribution behavior in Ultimatum and Dictator Games. For specification u_X, an agent can be voluntarily generous – offering more than the share $y^*(i_Y^{U\prime})$ needed to ensure acceptance – in the Ultimatum Game, while claiming the entire surplus in the Dictator Game. Algebraic manipulation of $u_X(0.5,0.5)>u_X(1-y^*(i_Y^{U\prime}),y^*(i_Y^{U\prime}))$ implies that the distributor prefers to voluntarily offer $y=0.5$ whenever

$$i_X^U > \hat{i}_X(i_Y^{U\prime}) \equiv \frac{1}{2}\sqrt{1+2i_Y^{U\prime 2}-2i_Y^{U\prime}\sqrt{i_Y^{U\prime 2}+1}}.$$

Hence,

$$y^*(i_X^U,i_Y^{U\prime}) = \begin{cases} i_Y^{U\prime}\sqrt{i_Y^{U\prime 2}+1}-i_Y^{U\prime 2}; & i_X^U \leq \hat{i}_X(i_Y^{U\prime}) \\ 0.5; & i_X^U > \hat{i}_X(i_Y^{U\prime}) \end{cases} \tag{8}$$

is X's equilibrium ultimatum offer to Y for preferences u_X.

For specification (2') of X's inequality aversion, i. e. utility function $\tilde{u}_X$, distributor X's optimal ultimatum offer is

$$\tilde{y}^*(i_X^U,i_Y^{U\prime}) = \begin{cases} \dfrac{i_X^U-1}{2i_X^U}; & i_Y^{U\prime} < \frac{i_X^U-1}{2\sqrt{i_X^U}} \\ i_Y^{U\prime}\sqrt{i_Y^{U\prime 2}+1}-i_Y^{U\prime 2}; & i_Y^{U\prime} \geq \frac{i_X^U-1}{2\sqrt{i_X^U}}. \end{cases} \tag{9}$$

The major difference between $y^*(i_X^U,i_Y^{U\prime})$ and $\tilde{y}^*(i_X^U,i_Y^{U\prime})$ is that the receiver's equality concern $i_Y^{U\prime}$ ceases to (positively) influence the offer $y^*(i_X^U,i_Y^{U\prime})$ above some threshold level; this is due to the 'bang-bang' character of the distributor's behavior. In contrast, greater $i_Y^{U\prime}$ keeps inducing a greater offer $\tilde{y}^*(i_X^U,i_Y^{U\prime})$.

For preference specification u_X, the average payoff to an agent with inequality aversion i^U interacting with an $i^{U\prime}$-type is[12]

$$\Pi_i^U(i^U,i^{U\prime}) = \begin{cases} \frac{1}{2}\left[1-i^{U\prime}\sqrt{i^{U\prime 2}+1}+i^{U\prime 2}+i^U\sqrt{i^{U2}+1}-i^{U2}\right]; & i^U \leq \hat{i}_X(i^{U\prime}) \wedge i^{U\prime} \leq \hat{i}_X(i^U) & \text{(I)} \\ \frac{1}{2}\left[1-i^{U\prime}\sqrt{i^{U\prime 2}+1}+i^{U\prime 2}+\frac{1}{2}\right]; & i^U \leq \hat{i}_X(i^{U\prime}) \wedge i^{U\prime} > \hat{i}_X(i^U) & \text{(II)} \\ \frac{1}{2}\left[\frac{1}{2}+i^U\sqrt{i^{U2}+1}-i^{U2}\right]; & i^U > \hat{i}_X(i^{U\prime}) \wedge i^{U\prime} \leq \hat{i}_X(i^U) & \text{(III)} \\ \frac{1}{2}; & i^U > \hat{i}_X(i^{U\prime}) \wedge i^{U\prime} > \hat{i}_X(i^U) & \text{(IV)} \end{cases} \tag{10}$$

For a parameter combination $(i^U,i^{U\prime})$ in the interior of region III, a mutant with $i^{U\prime\prime} \leq \hat{i}_X(i^{U\prime})$ fares better than the i^U-agent. Thus a parameter combination in

[12] For the case of $i^U \equiv i_X^U \equiv i_Y^U$, we at this point drop the redundant role subscript.

region I will be reached in the long run. A symmetric argument applies to region II. In region IV, the i^U-agent fares worse than a mutant with $i^{U\prime\prime} \leq \hat{i}_X(i^{U\prime})$. So, again, evolution will drive agents' inequality aversion via region II into region I. Within region I, however, the i^U-agent's payoff is increasing with i^U. So the unique equilibrium level of inequality aversion turns out to be the solution to $i^U = \hat{i}_X(i^U)$, i. e. fixed point $i^{U*} = 1/4 \cdot \sqrt{2} \approx 0.354$. This corresponds to a moderately equitable payoff distribution, $\pi_X = 3/4$ and $\pi_Y = 1/4$.

An even more equitable distribution results from our second specification of utility in the role of distributor, $\tilde{u}_X$. One obtains

$$\tilde{\Pi}_i^U(i^U, i^{U\prime}) = \begin{cases} \frac{1}{2}\left[\frac{i^U+1}{2i^U} + i^U\sqrt{{i^U}^2+1} - {i^U}^2\right]; & i^{U\prime} < \frac{i^U-1}{2\sqrt{i^U}} \wedge i^U \geq \frac{i^{U\prime}-1}{2\sqrt{i^{U\prime}}} \quad \text{(I)} \\ \frac{1}{2}\left[1 - i^{U\prime}\sqrt{{i^{U\prime}}^2+1} + {i^{U\prime}}^2 + \frac{i^{U\prime}-1}{2i^{U\prime}}\right]; & i^{U\prime} \geq \frac{i^U-1}{2\sqrt{i^U}} \wedge i^U < \frac{i^{U\prime}-1}{2\sqrt{i^{U\prime}}} \quad \text{(II)} \\ \frac{1}{2}\left[1 - i^{U\prime}\sqrt{{i^{U\prime}}^2+1} + {i^{U\prime}}^2 + i^U\sqrt{{i^U}^2+1} - {i^U}^2\right]; & i^{U\prime} \geq \frac{i^U-1}{2\sqrt{i^U}} \wedge i^U \geq \frac{i^{U\prime}-1}{2\sqrt{i^{U\prime}}} \quad \text{(III)} \end{cases} \tag{11}$$

In region I, the i^U-agent – voluntarily making generous ultimatum offers due to great inequality aversion – fares worse than a mutant with $i^{U\prime\prime} < (i^{U\prime}-1)/(2\sqrt{i^{U\prime}})$. So, there is downward pressure on i^U for parameter constellations in region I. However, once i^U and $i^{U\prime}$ are sufficiently similar, greater inequality aversion is beneficial: It implies getting a higher offer as receiver, while not affecting the respective agent's own distribution behavior. So evolution will lead to a monomorphism with unbounded inequality aversion, corresponding to equal splits in the Ultimatum Game ($\pi_X = \pi_Y = 1/2$).

On the one hand, greater concern for equality hurts a distributor when matched with someone whose inequality aversion is comparatively low and would therefore accept a smaller offer. On the other hand, greater observable inequality aversion helps a receiver. The latter advantage dominates the former disadvantage only when an agent has moderately low inequality concern if marginal disutility of inequality is decreasing. Above a critical level of parameter i, an agent would discontinuously offer fully equitable fifty-fifty splits as a distributor, but only get marginally better splits as a receiver. In contrast, if marginal disutility is increasing, receivers always get better treatment for greater i[13] while agents' distribution behavior becomes independent of their own type above a certain level i' in the population. In this case, the advantage of greater i persists and the disadvantage disappears.

So far, we have assumed that agents have an identical strength of inequality aversion in both roles X and Y, i. e. $i^U \equiv i_X^U \equiv i_Y^U$. What happens if, in contrast, i_X^U and i_Y^U may evolve *independently*? While it is always (weakly) better to have greater i_Y^U – inducing X to make a more equitable offer – it is (weakly) detrimental to have high i_X^U, i. e. to be voluntarily generous when drawn to act as the distributor.

[13] This is true for sufficiently pronounced inequality aversion, i. e. if $i \geq (i'-1)/(2\sqrt{i'})$.

So evolution will cause i_X^U to fall until it does no longer translate into positive offers. This endogenously yields one-sided inequality aversion, for which Güth and Napel (2003) find equal splits in the Ultimatum Game and complete exploitation in the Dictator Game regardless of the precise composition of the habitat. The analysis of the mixed habitat in the following section can thus be restricted to role-independent inequality aversion, i. e. $i^U \equiv i_X^U \equiv i_Y^U$ and $i^D \equiv i_X^D \equiv i_Y^D$.

4. EVOLUTION IN THE MIXED ENVIRONMENT

Now consider an environment or stylized 'game of life' that confronts agents with qualitatively different games. Our assumptions entail that agents correctly understand the (sub-)game they play, i. e. both players' feasible actions and subjective preferences. However, different cases can be distinguished concerning agents' intrinsic motivation in this wider context. Their personal attitude to inequality may be sensitive to the game form at hand. Alternatively, their intrinsic motivation in different games may be guided by the same moral views.

4.1. Perfect Discrimination

If agents are able to morally discriminate between different strategic situations, their inequality aversion can be conditioned on the game form they face and parameters i^U and i^D can evolve independently. Then the results obtained above can directly be transferred to the stylized 'game of life' comprising both Dictator and Ultimatum Game. Namely, parameter i^D will in equilibrium be below the critical level which induces equitable actions. Agents opportunistically maximize material reward in the Dictator Game – corresponding to the equilibrium division of surplus $(1, 0)$. In contrast, inequality aversion will be pronounced in the Ultimatum Game, implying equitable offers of half the surplus or at least a 75-25 division.

Since only parameter i_X had an impact in the Dictator Game and the distributor's strategic reaction to i_Y drove results for the Ultimatum Game, one can reinterpret i_X and i_Y as *role and game-independent* weights on advantageous and disadvantageous inequality, respectively (cf. the preferences studied by Fehr and Schmidt 1999). If these weights are allowed to evolve completely independently from each other, the concern for disadvantageous inequality will grow to significant levels while concern for advantageous inequality is unnoticeable.

4.2. No Discrimination

Being very inequality averse in the Ultimatum Game but not caring an iota about the surplus distribution in the Dictator Game is reminiscent of schizophrenia. It is plausible that agents' imperfect mental model of the world requires at least some link between the intrinsic motivation in both games. As a benchmark case, let

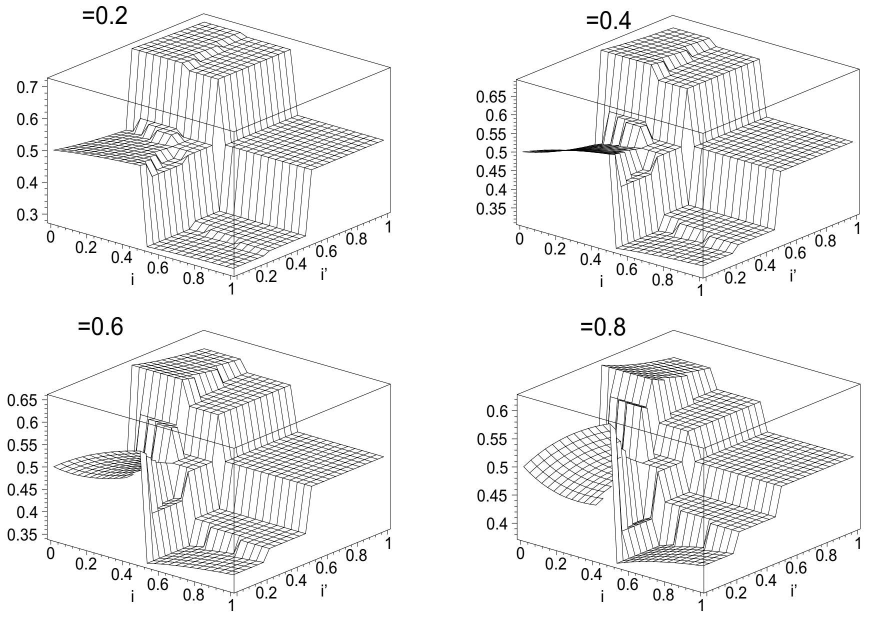

FIG. 1 Expected payoff to agent with aversion i for utility u_X

us investigate whether a noticeable level of inequality aversion – inducing deviations from materialistic optimization – will be observed if an agent's inequality aversion is universal, i. e. the same parameter $i \equiv i^U \equiv i^D$ applies to Dictator and Ultimatum Game.

In this situation, an i-agent's average payoff from playing the Ultimatum Game with probability λ and the Dictator Game with probability $1 - \lambda$ is

$$\Pi_i(i, i') = \lambda \Pi_i^U(i, i') + (1 - \lambda)\Pi_i^D(i, i') \tag{12}$$

or

$$\tilde{\Pi}_i(i, i') = \lambda \tilde{\Pi}_i^U(i, i') + (1 - \lambda)\tilde{\Pi}_i^D(i, i') \tag{13}$$

depending on the assumed form of inequality aversion. Figs. 1 and 2 depict corresponding payoffs in the mixed environment for several values of λ.

Though the payoff landscapes for $\Pi_i(i, i')$ shown in Fig. 1 are fairly complex, their evolutionary implications are straightforward: The Dictator Game only contributes evolutionary pressure towards the $i, i' \leq 0.5$ region, wherein all (i, i')-combinations receive identical Dictator Game payoffs. The only stable level of inequality aversion in the Ultimatum Game is $i^{U*} = 1/4 \cdot \sqrt{2} < 0.5$. It lies in the mentioned region and, hence, is also the only stable level for the mixed environment:

Proposition 1. *If agents' intrinsic motivation in Dictator Game and Ultimatum Game is constrained to be identical with decreasing marginal disutility from in-*

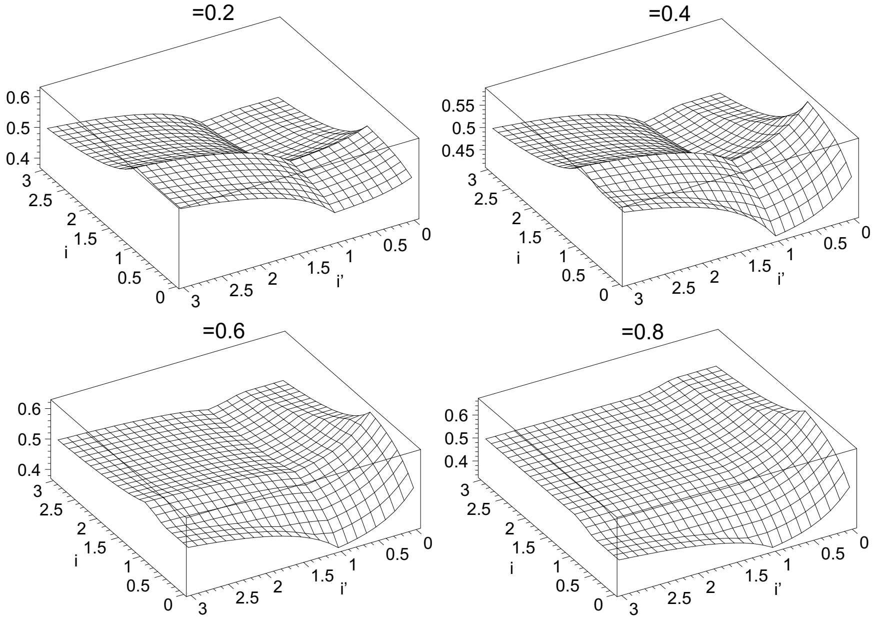

FIG. 2 Expected payoff to agent with aversion i for utility $\tilde{u}_X$

equality, evolution yields moderately equitable splits in the Ultimatum Game ($\pi_X = 3/4, \pi_Y = 1/4$*) and complete exploitation in the Dictator Game (*$1 - \pi_X = \pi_Y = 0$*).*

This result demonstrates that a mixed environment need not imply behavior different from that in isolated games. Though the presence of the Ultimatum Game fixes the hitherto rather indeterminate level of inequality aversion in the Dictator Game ($i \in [0, 0.5]$) to a particular one ($i = i^{U*}$), this does not affect offers.

For the alternative specification for role X which is underlying $\tilde{\Pi}_i(i, i')$, the situation is more complex. For all values of λ shown in Fig. 2, $(i, i') = (1, 1)$ is a saddle point which makes the long run level of inequality aversion $i = 1$, corresponding to the quite equitable surplus distribution $(2 - \sqrt{2}, \sqrt{2} - 1) \approx (0.59, 0.41)$ in the Ultimatum Game and $(1, 0)$ in the Dictator Game. This is quite well visualized by Fig. 2 for $\lambda = 0.2$ or 0.4, but much less so for $\lambda = 0.6$ or even $\lambda = 0.8$. In fact, this last share of the Ultimatum Game is close to a threshold which changes the long-run prediction. For $i, i' \geq 1$ we have

$$\frac{\partial \tilde{\Pi}_i(i, i')}{\partial i} = \frac{1}{2}\lambda \left[\sqrt{i^2 + 1} + \frac{i^2}{\sqrt{i^2 + 1}} - 2i\right] - (1 - \lambda)\frac{1}{4i^2}$$

as the i-agent's marginal payoff change if his or her inequality aversion increases. Equating this with zero, one obtains the following relationship between the Ultimatum Game's share in our stylized 'game of life', λ, and the long run level of $i > 1$

picked by evolution, corresponding to a saddle point of $\tilde{\Pi}_i(i, i')$:

$$\lambda = \frac{\sqrt{i^2+1}}{4i^4 + 2i^2 - 4i^3\sqrt{i^2+1} + \sqrt{i^2+1}}. \tag{14}$$

It turns out that $\bar{\lambda} = \sqrt{2}/(6 - 3\sqrt{2}) \approx 0.805$ is the critical level at which $i = 1$ ceases to be selected by evolution. For $\lambda > \bar{\lambda}$, a level $i > 1$ will result in the long run – implying offers greater than $y = \sqrt{2} - 1$ in the Ultimatum Game and also positive offers in the Dictator Game. As can be deduced from (14), the stable level of i is unbounded as $\lambda \to 1$, inducing equitable allocations in both Ultimatum and Dictator Games in the limit. Summing up we have:

Proposition 2. *If agents' intrinsic motivation in Dictator Game and Ultimatum Game is constrained to have identical strength and if in role X marginal disutility from inequality is increasing,*[14] *the Ultimatum Game's share λ determines which level of inequality aversion is selected. For $\lambda \leq \bar{\lambda}$, one observes a distribution close to* $(0.59, 0.41)$ *in the Ultimatum Game and complete exploitation in the Dictator Game. For $\lambda > \bar{\lambda}$, the distributions in the Ultimatum Game and the Dictator Game become strictly more equitable the larger is λ.*

For utility specification u_X, the discontinuity in optimal ultimatum offers tightly limits the evolutionary advantage of inequality aversion in the Ultimatum Game and thereby also in the mixed environment. In contrast, greater concern for equity increases fitness in ultimatum interactions for specification $\tilde{u}_X$ even if inequality aversion in the population is already high (though at a diminishing rate). The optimal dictator offer, $\tilde{y}^{**}$, represents the cost of inequality aversion; it is strictly concave and increasing for $i \geq 1$. If agents are dictators sufficiently often, the diminished marginal benefit of increasing i beyond $i = 1$ in the Ultimatum Game is outweighed by a big jump of marginal cost in the Dictator Game (from zero for $i < 1$ to 1/4 just above $i = 1$); i stays at the largest level compatible with zero dictator offers. However, if the Dictator Game is a comparatively rare event, benefit of higher i in the Ultimatum Game and cost in the Dictator Game balance at some $i > 1$. This corresponds to positive dictator offers and still somewhat inequitable ultimatum offers – none of which is observed when games are analyzed in isolation. A sufficiently great share of the Ultimatum Game imparts benevolence to dictators; a positive share of the Dictator Game restricts distributors' offers and receivers' acceptance thresholds in the ultimatum game.

[14] Note that in role Y marginal disutility is still decreasing. Quadratic utility in both roles X and Y would yield similar 'game of life' effects, albeit with higher threshold $\bar{\lambda}$ and a less equitable limit distribution because receivers would be less demanding in the parameter range that corresponds to strategic rather than voluntary ultimatum offers.

5. EXTENSIONS

Since dictators are only concerned with advantageous inequality and ultimatum receivers care only about disadvantageous inequality, game-specific parameters i^D and i^U could be interpreted as game-independent parameters i^+ and i^- measuring an agent's aversion to advantageous and disadvantageous inequality, respectively. Independent evolution of i^+ and i^- in a stylized 'game of life' consisting only of Ultimatum Game and Dictator Game would result in pronounced concern with disadvantageous inequality, but no (noticeable) aversion against advantageous inequality.

Is this finding robust if agents must simultaneously trade off *both* types of inequality and their material payoff? Such a three-way trade-off is e. g. required in the *3-Player Ultimatum Game*: Player X (proposer) suggests a division $(1-y-z, y, z)$ with $y+z \leq 1$ and $y, z \geq 0$ to player Y (responder), which upon acceptance by Y yields the payoff $(\pi_X, \pi_Y, \pi_Z) = (1-y-z, y, z)$ for X, Y, and a third player Z (dummy), respectively. Rejection by Y yields $(\pi_X, \pi_Y, \pi_Z) = (0,0,0)$.[15]

We concentrate on the case of *self-centered inequality aversion* (see Fehr and Schmidt 1999), i. e. a given player cares about being better or worse off than the two remaining players, but suffers no direct disutility from any inequality between the latter two. Moreover, we assume that advantageous (disadvantageous) inequality between player j and k and advantageous (disadvantageous) inequality between j and $l \neq k$ are *perfect substitutes* to player j.[16] For example, X would be indifferent between distribution $(0.5, 0.25, 0.25)$ and $(0.5, 0.4, 0.1)$. Considering the case of increasing marginal disutility in all roles, i. e. also as a responder, we have

$$u_j(\pi_X, \pi_Y, \pi_Z) = \pi_j - i^- \cdot \Big(\sum_{k \neq j} \max\{\pi_k - \pi_j, 0\}\Big)^2 - i^+ \cdot \Big(\sum_{k \neq j} \max\{\pi_j - \pi_k, 0\}\Big)^2$$

for a given agent in role $j = X, Y, Z$. These utility functions can be considerably simplified in our context (see Güth and Napel 2003, pp. 25f). Namely, we can restrict attention to payoffs $\pi_X \geq \pi_Y \geq \pi_Z$, and therefore use the following simplified utility functions:

$$\begin{aligned} u_X(\pi_X, \pi_Y, \pi_Z) &= \pi_X - \tfrac{i^+}{4} \left(2\pi_X - \pi_Y - \pi_Z\right)^2 \\ u_Y(\pi_X, \pi_Y, \pi_Z) &= \pi_Y - \tfrac{i^-}{4} \left(\pi_X - \pi_Y\right)^2 - \tfrac{i^+}{4} \left(\pi_Y - \pi_Z\right)^2, \end{aligned}$$

[15] Laboratory behavior in this game has been studied by Güth and Van Damme (1998), Güth, Schmidt, and Sutter (2003) as a newspaper experiment, and by Brandstätter and Güth (2002) as the last phase of a more complex experiment. Bolton and Ockenfels (1999) have tried to account by inequity aversion for the basic observation (of Güth and Van Damme) that the pie is essentially shared by proposer and responder only.

[16] This may be justified by noticing that it allows maximal flexibility in shifting inequality to the player with smallest power – possibly an evolutionary advantage.

where the scaling factor for i^- is chosen such that responder behavior for $i^+ = \pi_Z = 0$ is as it would be in the (2-player) Ultimatum Game; that for i^+ is taken to be the same.[17]

In the 3-Player Ultimatum Game, proposer X with parameter i_X^+ facing responder Y with parameters i_Y^- and i_Y^+ proposes the allocation $\pi = (1-y-z, y, z)$ solving

$$\begin{array}{ll} & \max\limits_{0 \le z \le y \le 1-y-z} (1-y-z) - \frac{i_X^+}{4}\left(2-3y-3z\right)^2 \\ \text{s.t.} & y - \frac{i_Y^-}{4}\left(1-2y-z\right)^2 - \frac{i_Y^+}{4}(y-z)^2 \ge 0. \end{array}$$

Player X is concerned with aggregate advantageous inequality $2\pi_X - \pi_Y - \pi_Z$, but not the distribution between Y and Z. If the responder's acceptance constraint is binding, any *given* amount $y + z \equiv p \in (0, 2/3]$ should therefore be distributed between Y and Z in the way which responder Y prefers most[18] (as long as it leaves X weakly better off than both Y and Z). We can thus replace z in above maximization problem by Y's (hypothetical) *dictator offer* $z^{**}(p, i_Y^-, i_Y^+)$ to Z, which is derived from

$$(*) \quad \max_{0 \le z \le p} (p-z) - \frac{i_Y^-}{4}\left((1-p)-(p-z)\right)^2 - \frac{i_Y^+}{4}\left((p-z)-z\right)^2.$$

Taking into account that Y is better off rejecting X's offer if it would yield negative utility, given parameters i_Y^- and i_Y^+ imply either of two optimal strategies for player Y (see the Appendix for details):

1. Offers $p < p_2(i_Y^-, i_Y^+)$ are rejected, offers $p_2(i_Y^-, i_Y^+) \le p \le p_1(i_Y^-, i_Y^+)$ are accepted and fully appropriated, and offers $p_1(i_Y^-, i_Y^+) < p$ are accepted and shared with player Z.

2. Offers $p < p_3(i_Y^-, i_Y^+)$ are rejected, and offers $p_3(i_Y^-, i_Y^+) \le p$ are accepted and shared.

If the allocation $(1-p, p - z^{**}(p, i_Y^-, i_Y^+), z^{**}(p, i_Y^-, i_Y^+)))$ with strictly positive p is proposed by player X for *strategic* reasons, the optimal offer involves

$$p^*(i_Y^-, i_Y^+) = \begin{cases} p_2(i_Y^-, i_Y^+); & p_3(i_Y^-, i_Y^+) \le p_1(i_Y^-, i_Y^+) \\ p_3(i_Y^-, i_Y^+); & p_3(i_Y^-, i_Y^+) > p_1(i_Y^-, i_Y^+) \end{cases} \tag{15}$$

and is accepted in equilibrium. Player Y's corresponding material payoff, $y^*(i_Y^-, i_Y^+) \equiv p^*(i_Y^-, i_Y^+) - z^{**}(p^*(i_Y^-, i_Y^+), i_Y^-, i_Y^+)$, is illustrated in Fig. 3.

For $p_3(i_Y^-, i_Y^+) \le p_1(i_Y^-, i_Y^+)$, implying that p is fully appropriated, player Y's material payoff increases in Y's aversion against *both* disadvantageous and advantageous inequality. A higher level of inequality aversion implies that a given material

[17] There are several alternatives to this. One is to divide total inequality of either type by the number $n-1$ of other players in the considered game, as proposed by Fehr and Schmidt (1999).

[18] This is reminiscent of labor division in controlling inequity: whereas X decides how much to give to the 'less powerful' (Y and Z), responder Y takes care of the 'powerless' (Z).

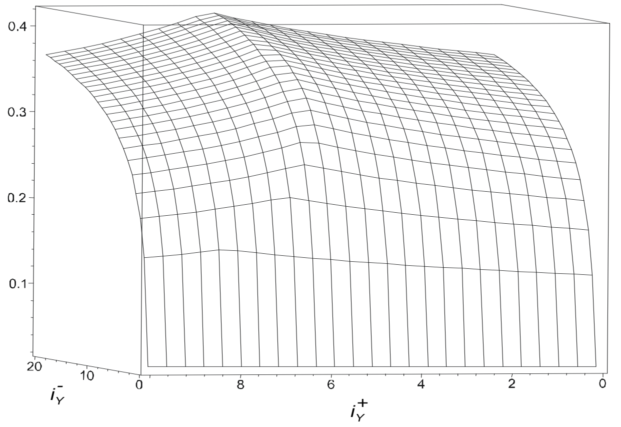

FIG. 3 Player Y's payoff in the 3-Player Ultimatum Game for sufficiently small i_X^+

payoff y^* yields smaller utility to player Y. Player X thus has to offer Y a greater material payoff to induce acceptance. This reduces disadvantageous inequality for Y, but ceteris paribus, i. e. for fixed $z = 0$, increases advantageous inequality – which again has to be compensated by material payoff, etc. The power of this beneficial multiplier effect increases with i_Y^+.

For $p_3(i_Y^-, i_Y^+) > p_1(i_Y^-, i_Y^+)$, implying that p is shared with dummy player Z, player Y's material payoff increases in Y's aversion against disadvantageous inequality, but *falls* with further aversion against advantageous inequality. The logic is similar to that of the appropriation regime except that it is now optimal for player X to deal with Y's utility losses also by increasing the proposed material payoff for player Z. Greater aversion against advantageous inequality thus results in a bigger total pie p^* of which, however, player Y keeps a *smaller* part than before.

Parameter i^- matters only in role Y. There is persistent upwards pressure on it, implying ever more equal shares for players X and Y. In contrast, the parameter i^+ measuring aversion against advantageous inequality has an impact on behavior in both roles X and Y. In role Y, it will be selected *for* if i^+ is small enough to make $z = 0$ the optimal assignment to dummy player Z: Moral suffering from positive i^+ in role Y is sufficient to prompt material compensation from player X, but is too weak to make the responder ask for a greater payoff for powerless dummy Z. Aversion against advantageous inequality will be selected *against* in role Y if i^+ is big enough to make player Y call for $z > 0$, i. e. when it translates into a contribution to inequality reduction by player Y. If parameter i^+ had no effect on behavior in role X, its stable level for given i^- would exactly be at the boundary between the two distribution regimes, defined by $p_3(i^-, i^+) = p_1(i^-, i^+)$ or

$$\hat{i}^+(i^-) \equiv \frac{12 + 20i^- - {i^-}^2 + (2 + i^-)\sqrt{({i^-}^2 + 84i^- + 36)}}{8i^-}. \tag{16}$$

This level approaches $\lim_{i^- \to \infty} \hat{i}^+(i^-) = 8$ as i^- is driven upwards.

However, in role X, high i^+ can imply evolutionary detrimental non-strategic generosity. We solve

$$\max_{0 \leq p \leq \frac{2}{3}} (1 - p) - \frac{i_X^+}{4} (2 - 3p)^2 ,$$

to identify which total offer p player X *voluntarily* (even as a dictator) would make to Y and Z. This yields

$$\hat{p}(i_X^+) = \begin{cases} 0; & i_X^+ \leq \frac{1}{3} \\ \frac{2}{3} - \frac{2}{9i_X^+}; & i_X^+ > \frac{1}{3}. \end{cases} \tag{17}$$

$i_X^+ = 4/3$ implies a voluntary offer of half the total surplus. Agents in a population with $i^+ > 4/3$ would offer more than is necessary to make them accept as responders for any $i^- \geq 0$, but then mutants with lower i^+ could successfully invade.

Therefore, if the 3-Player Ultimatum Game is played in isolation, i^- rises without bound and i^+ increases slowly enough such that agents neither make voluntarily generous proposals nor induce a positive share for player Z through their responder behavior.[19] Hence, the surplus distribution approaches $(1/2, 1/2, 0)$ and parameter i^+ rises to $4/3$. The payoff advantage of sacrifice-free concern for player Z's lot vanishes as i^- reaches ever higher levels.

When the 3-Player Ultimatum Game is added to the multi-game environment, i.e., enriches the 'game of life', the Ultimatum Game has no qualitative effect on the evolutionary pressures on i^+ and i^- induced by the 3-Player Ultimatum Game: Greater i^- is beneficial, while any i^+ that does not lead to a voluntarily generous proposal yields identical fitness. The Dictator Game does not at all affect agents' fitness from i^-. Therefore evolution in a stylized 'game of life' comprising all three types of interaction must bring about equal splits between players X and Y as i^- approaches infinity. In the Dictator Game, agents with $i^+ > 1$ make offers that are strictly positive (and continuously increasing to 0.5 for $i^+ \to \infty$). For any given share $\mu > 0$ of the Dictator Game in agents' environment there is therefore a fixed fitness cost born by an agent with $i^{+\prime} > 1$ compared to an agent with $i^+ = 1$. Since any fitness benefit of $i^{+\prime} \in (1, 4/3]$ in the 3-Player Ultimatum Game vanishes as $i^- \to \infty$, $i^+ = 1$ must prevail in the long run.

In summary, a stylized 'game of life' which comprises Dictator Game, Ultimatum Game, and 3-Player Ultimatum Game, in which agents face no exogenous restriction

[19] The former constraint, formally $\hat{p}(i^+) \leq p_2(i^-, i^+)$, is the more demanding one here. This is sensitive to the scaling factor for inequality aversion chosen in u_X.

on morally distinguishing between advantageous and disadvantageous inequality, would bring about agents that are highly sensitive to disadvantageous inequality but only minimally concerned with being better off than others – just as in the two-game habitat. In the long run, surplus is fully appropriated in Dictator Games by the proposer and in (2 or 3-Player) Ultimatum Games by the proposer and responder. The latter two share surplus equally for strategic reasons, while the powerless third player receives nothing. Presence of the dummy player nevertheless has an impact in the short and medium run: A proposer – for above utility specification u_X – is more likely to be voluntarily generous in the light of two rather than only one worse-off agents, and a responder, interestingly, can credibly ask for an extra share in view of the emotional costs of being better off than someone else.

Another modification is analyzed in detail by Berninghaus, Korth, and Napel (2003) who replace purely outcome-oriented concern for distributional equity by a preference for *intention-based* reciprocal behavior. In the tradition of psychological game theory (Geanakoplos, Pearce, and Stacchetti 1989), Berninghaus et al. consider agents who have the reciprocal preferences specified by Falk and Fischbacher (2001) and use implied equilibrium behavior to determine material payoffs of different preference types. Their long-run predictions are qualitatively similar to the case of increasing marginal disutility of inequality considered above.

6. CONCLUSIONS

Our main aim has been to point out and overcome a shortcoming of most evolutionary game theory (see Hammerstein and Selten 1994, Weibull 1995, and Samuelson 1997 for surveys), namely studying evolution of behavior and its underlying preferences for just one (often numerically specific) game. There are a few basic traits like inequality aversion, reciprocity, truthfulness, trustworthiness, and their respective counterparts which seem to structure our decision behavior in a large, possibly infinite number of decision environments. It is of utmost importance to study the evolution of such general characteristics for habitats comprising strategically different games. The goal is to capture at least some aspects of our complex 'game of life' and of how we manage it by relying on several fundamental behavioral dispositions.

Of course, we do not (and probably will never) come close to an adequate model of the 'game of life'. Our attempt has concentrated on combinations of the Ultimatum Game, a paradigm of close strategic interaction as in private affairs, and the Dictator Game without strategic interdependencies. This is a first step that already yields, in our view, interesting results when agents' possibility of developing game-specific or role-specific preferences is restricted.

Our analysis concentrated on the case of commonly known intrinsic motivation which does not only affect own but also others' behavior. Clearly, private information about moral concerns without any possibility of correct observation or credible signaling would change our conclusions. But to have information about one's opponent at least occasionally does not seem unrealistic to us.

Depending on the relative weights of the Ultimatum and the Dictator Game, and also the preference specification (increasing vs. decreasing marginal disutility of inequality) one will observe either universal equity or (partial) exploitation of recipients by distributors. Presence of the Dictator Game may restrict offers in the Ultimatum Game, while high frequency of the latter can impart benevolence on dictators. Such results reveal that evolutionary studies for structurally richer habitats can yield much more interesting and intuitive results than game-specific evolution. It is noteworthy in this context that the earlier-mentioned conjecture that evolutionary benefits of non-individualistic preferences would erode as their domain is extended or as one moves from perfect to imperfect moral discrimination of games, is only weakly confirmed: Inequality aversion that is sufficiently strong to be noticed in the Ultimatum Game seems a fairly robust feature.

The price of considering richer habitats is, of course, the increase in complexity and a need for richer case distinctions. Here, empirical studies could help, e. g. in the sense of indicating limits for the relevant range of parameters, like λ or μ which measure the importance of different strategic aspects of the 'game of life'. Though one will remain far from studying the actual 'game of life', we should approach it more closely in order to give a satisfying answer to the question of why mankind has developed a capability of empathy and reciprocity.

APPENDIX – Y'S OPTIMAL STRATEGY

Solving $(*)$ yields

$$z^{**}(p, i_Y^-, i_Y^+) = \begin{cases} 0; & i_Y^+ \leq \frac{2+i_Y^-(1-2p)}{2p} \\ \dfrac{2p(i_Y^- + i_Y^+) - i_Y^- - 2}{i_Y^- + 4i_Y^+}; & i_Y^+ > \frac{2+i_Y^-(1-2p)}{2p}. \end{cases}$$

Player Y compares this optimal distribution of pie p, which Y and Z can share, to allocation $(0, 0, 0)$ which would result from rejecting. For the nontrivial case $i_Y^- + i_Y^+ > 0$ consider first $z^{**}(p, i_Y^-, i_Y^+) = 0$, i. e.

$$i_Y^+ \leq \frac{2 + i_Y^-(1-2p)}{2p} \Longleftrightarrow p \leq \frac{2 + i_Y^-}{2(i_Y^- + i_Y^+)} \equiv p_1(i_Y^-, i_Y^+). \tag{18}$$

$p_1(i_Y^-, i_Y^+)$ denotes the maximal level of p such that *appropriating pie p is weakly preferred by player Y to sharing it.* In this case, the individual rationality constraint $u_Y(p, z^{**}) \geq 0$ amounts to

$$i_Y^+ \leq \frac{4p - i_Y^-(1-2p)^2}{p^2}.$$

Given $i_Y^- + i_Y^+ > 0$ and (18), this is equivalent to

$$p_2(i_Y^-, i_Y^+) \equiv \frac{2(i_Y^- + 1) - \sqrt{4 + 8i_Y^- - i_Y^- i_Y^+}}{4i_Y^- + i_Y^+} \leq p.$$

$p_2(i_Y^-, i_Y^+)$ denotes the minimal level of p such that *appropriating pie p yields non-negative utility* to player Y; this does not entail optimality of $z = 0$. Player Y thus finds it optimal to accept a total pie p and to appropriate it completely if

$$p_2(i_Y^-, i_Y^+) \leq p \leq p_1(i_Y^-, i_Y^+). \tag{19}$$

It can happen that the minimal level of p such that appropriation is individually rational exceeds the level of p such that appropriation is preferred to sharing, i. e. $p_2(i_Y^-, i_Y^+) > p_1(i_Y^-, i_Y^+)$. In this case, (19) cannot be satisfied and any acceptable offer by player X must involve strictly positive material payoffs for Y *and* Z.

Second, consider $z^{**}(p, i_Y^-, i_Y^+) > 0$, corresponding to

$$p > p_1(i_Y^-, i_Y^+). \tag{20}$$

In this case, the individual rationality constraint $u_Y(p, z^{**}) \geq 0$ amounts to

$$i_Y^+ \leq \frac{4 + 4i_Y^-(1-p)}{i_Y^-(3p-2)^2 - 8p} \text{ if } i_Y^- > \frac{8p}{(3p-2)^2}. \tag{21}$$

If $i_Y^- = 0$, player Y's individual rationality imposes no restriction on i_Y^+ or p. For $i_Y^+ = 0$, above condition is always satisfied. For $i_Y^- i_Y^+ > 0$, (21) is equivalent to

$$p_3(i_Y^-, i_Y^+) \equiv \frac{-2i_Y^- + 6i_Y^- i_Y^+ + 4i_Y^+ - 2\sqrt{(4i_Y^+ + i_Y^-)(i_Y^- + 3i_Y^- i_Y^+ + i_Y^+)}}{9i_Y^- i_Y^+} \leq p$$
$$\text{if } p \leq \frac{6i_Y^- + 4 - 4\sqrt{3i_Y^- + 1}}{9i_Y^-}. \tag{22}$$

Whenever p is below $p_3(i_Y^-, i_Y^+)$, the qualifying if-statement is true. Therefore, the latter can be dropped. Combining (20) and (22), a total offer p to players Y and Z will be accepted and leads to a positive 'dictator offer' z^{**} whenever

$$\max\left\{p_1(i_Y^-, i_Y^+), p_3(i_Y^-, i_Y^+)\right\} \leq p. \tag{23}$$

$p_3(i_Y^-, i_Y^+)$ denotes the minimal level of p such that *optimal 'unconstrained sharing' of pie p* – ignoring a possible negativity of $z^{**} = (2p(i_Y^- + i_Y^+) - i_Y^- - 2)/(i_Y^- + 4i_Y^+)$ – *yields non-negative utility* to player Y; $p_1(i_Y^-, i_Y^+)$ ensures $z^{**} > 0$.

It can be checked that $p_2(i_Y^-, i_Y^+) \geq p_3(i_Y^-, i_Y^+)$, i. e. the minimal p such that appropriation ($z = 0$) is individually rational is always weakly larger than the p which makes optimal 'unconstrained sharing' (possibly involving a negative allocation for player Z) individually rational.[20] Therefore $p_3(i_Y^-, i_Y^+) > p_1(i_Y^-, i_Y^+)$ implies

$p_2(i_Y^-, i_Y^+) > p_1(i_Y^-, i_Y^+)$, the condition which makes (19) impossible to be satisfied. Optimal responder behavior is thus well-defined.[21]

REFERENCES

Berninghaus, S. K., C. Korth, and S. Napel (2003). Reciprocity – An indirect evolutionary analysis. SFB 504 Discussion Paper 03-32, University of Mannheim.

Binmore, K. G. (1994). *Game Theory and the Social Contract – Volume I: Playing Fair.* Cambridge, MA: MIT Press.

Binmore, K. G. (1998). *Game Theory and the Social Contract – Volume II: Just Playing.* Cambridge, MA: MIT Press.

Bolton, G. E. (1991). A comparative model of bargaining: Theory and evidence. *American Economic Review 81*(5), 1096–1136.

Bolton, G. E. and A. Ockenfels (1999). An ERC-analysis of the Güth-van Damme game. *Journal of Mathematical Psychology 42*(2), 215–226.

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

Brandstätter, H. and W. Güth (2002). Personality in dictator and ultimatum games. *Central European Journal of Operations Research 3*(10), 191–215.

De Waal, F. (1998). *Chimpanzee Politics: Power and Sex Among Apes* (Revised ed.). Washington, DC: John Hopkins University Press.

Dekel, E., J. C. Ely, and O. Yilankaya (1998). Evolution of preferences. Mimeo, Tel Aviv University and Northwestern University.

Ely, J. C. and O. Yilankaya (2001). Nash equilibrium and the evolution of preferences. *Journal of Economic Theory 97*(2), 255–272.

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

Fehr, E. and K. M. Schmidt (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics 114*(3), 817–868.

Gardner, M. (1970). Mathematical games: The fantastic combinations of John Conway's new solitaire game 'life'. *Scientific American 223*(10), 120–123.

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

[20] Both coincide when $z^{**} = 0$ is the optimal 'unconstrained share', i. e. $p_2(i_Y^-, i_Y^+) = p_3(i_Y^-, i_Y^+)$ iff $p_2(i_Y^-, i_Y^+) = p_1(i_Y^-, i_Y^+)$. This implies a continuous transition between the two optimal strategy regimes and a continuous optimal offer $p^*(i_Y^-, i_Y^+)$ by player X.

[21] It is also implied that responder behavior is monotonic, i. e. if offer p is accepted by Y (and either appropriated or shared), so will any offer $p' > p$.

Güth, W. (1995). An evolutionary approach to explaining cooperative behavior by reciprocal incentives. *International Journal of Game Theory 24*(4), 323–344.

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: Experimente in der Ökonomik*, pp. 113–139. Marburg: Metropolis-Verlag.

Güth, W. and S. Napel (2003). Inequality aversion in a variety of games – An indirect evolutionary analysis. IAW Discussion Paper 133, Dept. of Economics, University of Hamburg.

Güth, W. and B. Peleg (2001). When will payoff maximization survive? *Journal of Evolutionary Economics 11*(5), 479–499.

Güth, W., C. Schmidt, and M. Sutter (2003). Fairness in the mail and opportunism in the internet – A newspaper experiment on ultimatum bargaining. *German Economic Review 4*(2), 243–265.

Güth, W. and E. Van Damme (1998). Information, strategic behavior, and fairness in ultimatum bargaining: An experimental study. *Journal of Mathematical Psychology 42*(2-3), 227–247.

Hammerstein, P. and R. Selten (1994). Game theory and evolutionary biology. In R. J. Aumann and S. Hart (Eds.), *Handbook of Game Theory*, Volume II, pp. 929–993. Amsterdam: North-Holland.

Huck, S. and J. Oechssler (1999). The indirect evolutionary approach to explaining fair allocations. *Games and Economic Behavior 28*(1), 13–24.

Kirchsteiger, G. (1994). The role of envy in ultimatum games. *Journal of Economic Behavior and Organization 25*(3), 373–389.

Koçkesen, L., E. A. Ok, and R. Sethi (2000a). Evolution of interdependent preferences in aggregative games. *Games and Economic Behavior 31*(2), 303–310.

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

Ok, E. A. and F. Vega-Redondo (2001). On the evolution of individualistic preferences: An incomplete information scenario. *Journal of Economic Theory 97*(2), 231–254.

Possajennikov, A. (2000). On the evolutionary stability of altruistic and spiteful preferences. *Journal of Economic Behavior and Organization 42*(1), 125–129.

Poulsen, A. and O. Poulsen (2005). Endogenous preferences and social institutions. Working paper, Dept. of Economics, Aarhus School of Business.

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

Samuelson, L. (1997). *Evolutionary Games and Equilibrium Selection*. Cambridge, MA: MIT Press.

Samuelson, L. (2001). Introduction to the evolution of preferences. *Journal of Economic Theory 97*(2), 225–230.

Sethi, R. and E. Somanathan (2001). Preference evolution and reciprocity. *Journal of Economic Theory 97*(2), 273–297.

Sigmund, K. (1995). *Games of Life: Explorations in Ecology, Evolution, and Behaviour*. London: Penguin.

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