Playing Fair:
Rationality and Norm-guided Behavior in Games

Markus Pasche
University of Jena, Faculty of Economics
Carl-Zeiss-Str. 3, D-07743 Jena, email: m.pasche@wiwi.uni-jena.de

Abstract: There is robust experimental evidence that in the ultimatum game real players often prefer a fair allocation which seems to be in contrast to rational decision making. In this paper players can be committed to rational maximizing behavior or to norm-guided fair behavior as two possible behavioral rules. It is argued that behavioral rules are adopted according to their expected success before the ultimatum game is conducted with randomly chosen players. Using the concept of behavioral equilibrium profiles it is shown that conditional to the information status the players may adopt the fair behavioral rule instead of maximizing. Furthermore, conditions are derived where maximizing and fair behavior are both parts of a behavioral equilibrium profile. Also the relation to the indirect evolutionary approach is briefly discussed.

Keywords: Rationality, fairness, ultimatum game, behavioral equilibrium

JEL-Classification: C70, C72, C78, D63

1 Introduction: behavioral explanations in game theory

The ultimatum game is considered in its basic two-person form: A player X (proposer) has to make a proposal how to distribute a given amount of money M to both players $(\delta M, (1-\delta)M),\ \delta \in [0,1]$. The player Y (responder) can accept or reject the proposal. In case of acceptance the players receive a payoff according to the proposed allocation, in case of rejection both players get nothing. This game has been studied in various variants, and experiments have been conducted under different conditions. This paper do not aim to give an overview about the experimental results (cf. Camerer/Thaler 1995, Güth 1995, Huck 1997). One main result is that a significant part of real players (proposer as well as responder) seem to be intrinsically motivated to implement a fair allocation. This is in contrast to rational behavior which predicts an allocation $(M-\epsilon, \epsilon),\ \epsilon \to 0$, since a rational responder will prefer a small positive outcome ϵ to a zero payoff. A rational proposer will anticipate this behavior and maximizes his own payoff by proposing the mentioned "unfair" allocation. In this paper, the terms rational, maximizing, and opportunistic are used as synonyms.

The attitude to play fair or to follow other social norms deserves a theoretical explanation. There exists a couple of concepts in the literature how to deal with these behavioral effects. One approach is to modify the utility function by adding new arguments in order to account for unequity-aversion, reciprocity or other motivational dispositions (see e.g. Rabin 1993, Falk/Fischbacher 1998, Bolton/Ockenfels 1999 among others). Although this is a straightforward way to fit the economic models with the observed data, it seems to be ad hoc as long as there is no economic explanation for the suggested specific preferential or motivational structures. It is not sufficient to claim that the additional factors refer to some psychological theory. Since the logic of the model implies that agents behave kind in some sense because they *prefer* kindness (in terms of utility), it is a usual neoclassical-type explanation with exception for some additional reasoning about the utility function. One may say that agents with social preferences are still in some sense maximizers. This way of explanation rarely fails because preferences, attitudes, and motivation are not directly observable states, hence every observed behavioral pattern can be "explained" by preference functions with certain additional arguments.

A more promising concept is the indirect evolutionary approach (see Güth/Kliemt 1998 and Güth/Pull 2002 for an application). This approach accounts for both, opportunistic rational behavior as well as intrinsically motivated or norm-guided behavior. The more or less strategic deliberation of choice is combined with evolutionary adaption according to the expected objective outcome. Agents may make decisions which seem to them subjectively preferable even if the objective outcome is not. Of course the evolutionary process will favor behavioral patterns with a high objective performance and rule out others. It may turn out that opportunistic behavior is not a dominant pattern, and that agents who follow certain social norms (captured in their utility function) may survive. In contrast to theories with ad hoc assumptions about social preferences the indirect evolutionary approach serves as an *explanation* for the evolution of preferential structures which induce non-opportunistic behavioral patterns.

In some sense the argumentation in this paper is similar to the indirect evolutionary approach but some shortcomings are avoided. First, there remain some methodological doubts whether different observable behavior can and should be explained by different non-observable states like preferences, motivation or attitudes which are expressed in terms of a modified utility function. Even if the emergence and stability of social preferences have been fully explained by an evolutionary model, the actual different (fair, unfair) behavior of two participients in an experiment must then be explained by their different, but unfortunately unobservable preferences. This is not fully satisfying. Further problems arise when the same players participate in a dictator game experiment and exhibit significant less fair behavior (cf. Forsythe et al. 1994). If it is reasonable to assume that preferences are manifest and long-lasting structures, such results breed the neccessity to construct more and more complicated utility functions and to show their evolutionary success. Some authors refer to Becker (1976) who had outlined the idea of evolution of (social) preferences (e.g. Huck 1997). In contrast, Becker also strongly doubts whether behavioral differences can and should be explained by preferential differences (Becker/Stigler 1977).

For an evolutionary explanation it is sufficient to show that *behavioral patterns* have a different performance and are hence either propagated or ruled out by the selection process. There is no need to argue that the underlying *preferences* are selected, even though this might actually be the case. Moreover, the concept of evolution of preferences implies a strict distinction between objective outcome (as the selection criterion for the evolutionary process) and subjective utility (which is the criterion for individual decision making). Just this is the second shortcoming of the approach. For an evolutionary theory in a Darwinian tradition it is neccessary to compare the different success of behavioral patterns by an *objective* measure. In a biological context this may be the reproductive power, and there are *external* factors which "decide" about survival and reproduction. The individuals are not aware whether their genotype is selected or not. In a social context the evolutionary mechanism is often semantically interpreted as adaption, learning, or imitation (because agents do not "die" or "reproduce" in an economic sense). The difference between outcome and utility then seems to be at least a doubtful concept for logical reasons because a decision making process is based on the individually perceived value of the consequences of decisions. In contrast to the biological context, *agents* adopt or change a behavioral rule, imitate other agents, learn something and adapt themselves to the environmental conditions. It is by no means clear why these agent based processes should be driven by objective outcomes of their decisions. Consider, for example, two agents A and B. The behavior of A leads to a distribution of profits with a high average profit but also a high varicance. Agent B´s behavior leads to a lower average profit but with a far less variance.

Now consider that both agents are risk averse. The distribution for B has a higher utility and is prefered by *both* agents. Hence, it would be rational for A to imitate (or adapt to) the behavior of B. Is A or B more "successful"? Which behavior "performs" better? The evolutionary selection mechanism refers to the objective (expected) outcome and selects agent A´s decision rule while B is outperformed.

In competitive markets – the predominant example to illustrate the idea of economic evolution – the notion of "evolutionary forces" is persuasive. The survival of firms (and their strategies) may be more or less correlated with their objective profits. But also in this competitive world it is not clear why firms with a profit margin lower than average (but eventually also with a lower variance) will be ruled out from the market. It has to be pointed out that for many other economic problems it is far less evident to argue with an external evolutionary pressure, e.g. consumer choice, games with non-monetary outcomes, games where the outcome is a multidimensional vector and so on. Also in ultimatum games it is less convincing to invoke terms like "competition". For what reasons should a certain behavioral rule which leads to shaes less than 50% of M but a high level of subjective welfare be outperformed? By what external forces? The evolution of preferences is therefore a type of explanation which may be applied very carefully in a limited set of cases.

In this paper we use the approach of behavioral equilibria and behavioral equilibrium profiles (cf. Pasche 2001). The argumentation is as follows: Rational maximizing decision making as well as adopting other arbitrary (e.g. norm-guided) patterns of behavior are interpreted as *behavioral rules* which are not a priori presumed but have to be explained. The adoption of a rule may be interpreted as a result of an individual decision or a learning or adaption process. Like in the indirect evolutionary approach this adoption is guided by the (expected) success, but in terms of the agent´s own utilities.

More formally, let S_i be the strategy space of agent $i = 1, ..., n$. A behavioral rule is then a map $f_i : \times_{j \neq i} S_j \rightarrow S_i$. Hence, $s_i \in f_i(s^e_{-i})$ (with $s^e_{-i} = (s^e_1, ..., s^e_{i-1}, s^e_{i+1}, ..., s^e_n)$) denotes a certain strategy of player i which is in acordance with the behavioral rule f_i given i´s beliefs regarding the strategies of the other players s^e_{-i}. This decision need not be a best response to the expected strategies, i.e. it need not maximize the utility $u_i(s_i, s^e_{-i})$. Nevertheless also maximizing behavior $s^{max}_i \in f^{max}_i(s^e_{-i}) = \arg\max_{s_i} u_i(s_i, s^e_{-i})$ is a special type of a behavioral rule. In an equilibrium the expected strategy choices have to be consistent with the realized choices so that $s_i \in f_i(s_{-i})$ holds true for all $i = 1, ..., n$. We call (s_i, s_{-i})

a *behavioral equilibrium* since all players choose their strategies in accordance with their adopted behavioral dispositions and have therefore no reason to change unilaterally the strategy. Hence, for maximizing behavior $(s_i^{max}, s_{-i}^{max})$ is the Nash equilibrium which is a special case of an behavioral equilibrium. Since the chosen strategies depend on the rules f_i we say that an behavioral equilibrium – if it exists – is *induced* by the vector $(f_1, ..., f_n)$. Let Ω denote the set of alternative behavioral rules, then $(f_1, ..., f_n) = (f_i, f_{-i})$ is called a *behavioral equilibrium profile* if no player i can benefit from changing unilaterally the rule f_i (comparing the payoffs in the induced equilibria). The adoption of a certain rule is then explained by being part of a behavioral equilibrium profile since a rule learning or adaption process cannot lead to a better performance anymore. In disequilibrium there is always the chance of discovering a better performing rule which may yield different outcomes. We apply this concept to the ultimatum game to analyse the conditions (a) for the occurence of fairness dispositions, (b) the simultanous occurance of fair and unfair behavior in the population.

2 The Ultimatum Game with Different Behavioral Rules

2.1 Assumptions

Consider a population of n players where two agents are randomly drawn to play the ultimatum game. At the first stage the players have to select a certain behavioral rule. At the second stage the nature decides randomly which player is the proposer and which one is the responder. Then the ultimatum game is conducted. The logic behind this structure is that social norms are assumed to be manifest behavioral dispositions – at least for a certain time. Such a behavioral disposition should be beneficial in numerous similar decision situations where the agent probably faces different roles: Sometimes he is the proposer (X), sometimes he is the responder (Y). It does not make sense to say that an agent has internalized a fairness norm but is deciding according to this norm only in cases where he is the responder. The probability to be in position X or Y are p_x and p_y and for simplicity it is assumed $p_x = p_y = 1/2$. The set Ω of behavioral rules contains the maximizing opportunistic rule O and the fairness rule F. For simplicity the monetary amount to allocate is normed to $M = 1$.

Opportunistic rule O:

Position X: Depending on the (beliefs about the) opponent's rule choose a proposal that maximizes the monetary outcome for X.

Position Y: Accept every proposal with a positive share for Y, otherwise be indifferent.

Fairness rule F:

Position X: Choose the proposal $(1/2+\phi, 1/2-\phi)$.

Position Y: Accept each proposal with a share for Y which is at least $1/2-\phi$, otherwise reject.

It is $\phi = \alpha(\frac{1}{2}-\epsilon), \alpha \in [0,1]$, so that α describes the *degree of fairness* ($\alpha = 0$ leads to equity and denotes therefore complete fairness, $\alpha = 1$ leads to the same allocation like an opportunistic player would propose, that means no fairness). The degree of fairness α is assumed to be exogeneously given and it is Common Knowledge.
Let μ be the share of agents in the population which have adopted the opportunistic rule, and μ is also Common Knowledge. Consider a utility function $u(z) = z^m$ with $0 < m < 1$ (risk aversion) and z as the outcome. To keep notation simple let $OX = u(1-\epsilon)$ and $OY = u(\epsilon)$ be the utilities of an unfair allocation, and $FX = u(1/2+\phi)$ and $FY = u(1/2-\phi)$ as the utilities of the fair (of degree α) allocation.
Now the ultimatum game is conducted on the last stage according to the adopted rules. The rules are either observable or the player have to build expectations on the opponent´s rule. Since the beliefs should be consistent in case of unobservable rules we assume that agents believe that the opponent follows the opportunistic rule O with probability μ, and the fairness rule F with probability $(1-\mu)$. Now at the first stage it is possible to calculate the expected utility for both rules. In an equilibrium profile each player has adopted the behavioral rule with the highest expected utility given the rules of the other players. We do not reason about how an equilibrium profile is constituted, e.g. by strategic considerations, learning or an adaption process. Nevertheless we will say that an player has an "incentive" to adopt a rule. This indicates that the approach follows Rubinstein (1998, 4) since the agents deliberate in some way *how* they make decisions. This deliberation is driven by subjective valuation of the outcomes.

2.2 Interactions with observable rules

First, consider a situation where the players can observe the adopted rules. It can be argued that in cases where the player know each other well like in a family or other social reference groups the adopted norms, attitudes or customs are Common Knowledge. We say that an agent who faces a well known responder is interacting in a "local group". Otherwise we talk about "anonymous interactions" where the opponent´s rule is private information. In a local group an opportunistic proposer will anticipate that a fair responder will reject an unfair allocation. Hence it is rational to propose the fair allocation (FX, FY) in case that the rational player is selected from nature to be in position X. Therefore the expected utilities $E[u_O]$ and $E[u_F]$ of the behavioral rules O and F are

$$\begin{aligned} E[u_O] &= p_x(\mu XO + (1-\mu)XF) + (1-p_x)(\mu YO + (1-\mu)YF) \\ &= \frac{1}{2}\mu((1-\epsilon)^m + \epsilon^m) + \frac{1}{2}(1-\mu)\left(\left(\frac{1}{2}+\phi\right)^m + \left(\frac{1}{2}-\phi\right)^m\right), \\ E[u_F] &= p_x XF + (1-p_x)YF \\ &= \frac{1}{2}\left(\frac{1}{2}+\phi\right)^m + \frac{1}{2}\left(\frac{1}{2}-\phi\right)^m. \end{aligned}$$

There is an incentive to select the fair rule if

$$E[u_F] - E[u_O] = -\frac{1}{2}\mu((1-\epsilon)^m + \epsilon^m) + \frac{1}{2}\mu\left(\left(\frac{1}{2}+\phi\right)^m + \left(\frac{1}{2}-\phi\right)^m\right) \geq 0$$

holds true, which is always the case for $0 < m < 1$. When interacting within a local group there is always an incentive to adopt the fair rule. The incentive is higher with a higher degree of fairness (lower α). The explanation is simple: Due to risk aversion a fair rule guarantees a relative "smooth" payoff, because also in the responder position the agent will receive a fair share, while the opportunistic agent will get sometimes very much and sometimes very low, if he meets an opportunistic opponent. This explanation is not neccessarily in contrast to the assumption of moral attidues of equity, internalized fairness norms, and empathy with other players. Moreover, these norms and attitudes may have been evolved *because* it is beneficial to reduce the high risk which would be the result of a pure opportunistic population.

Since the fair rule outperforms opportunism independently from μ there is only one behavioral equilibrium profile $(F, ..., F)$ or $\mu = 0$ respectively. In this model it is much harder to explain why player behave unfair rather than fair. This result is due to the strong information assumption and seems not to be realistic.

2.3 Anonymous interactions

The responder´s rule adopted in the first stage is considered to be private information. Since the players are randomly drawn from the population the proposer expects that the responder has adopted the opportunistic rule O with probability μ. Depending on this expectation an opportunistic player has to decide whether he proposes a fair or an unfair allocation in case of being in position X. Obviously with a high μ he will face with a high probability a maximizing opponent, hence he will choose the unfair proposal (OX, OY). If μ is sufficiently low there is a high risk that the responder is a fair player who will reject the allocation, hence the proposer chooses (FX, FY). In the latter case we call the behavior "imitating" because the player behaves like a fair agent, but only for opportunistic reasons. In the responder position, however, he will accept both, OY and FY.

In order to calculate the expected utilities of the rules the player have to build expectations about the share ν of opportunistic players who make unfair proposals, and $(1-\nu)$ of opportunistic players who imitate. Like it was argued for the expectation μ, we consider that the expected and the realized share of unfair opportunistic players ν are the same. The expected payoffs are now (u_{OI} for imitating opportunistic agents):

$$
\begin{aligned}
E[u_O] &= p_x(\mu XO + (1-\mu)\cdot 0) + (1-p_x)(\mu\nu YO + \mu(1-\nu)YF + (1-\mu)YF) \\
&= \frac{1}{2}\mu((1-\epsilon)^m + \nu\epsilon^m) + \frac{1}{2}(\mu(1-\nu) + (1-\mu))\left(\frac{1}{2}-\phi\right)^m, \\
E[u_{OI}] &= p_x XF + (1-p_x)(\mu\nu YO + \mu(1-\nu)YF + (1-\mu)YF) \\
&= \frac{1}{2}\left(\frac{1}{2}+\phi\right)^m + \frac{1}{2}\mu\nu\epsilon^m + \frac{1}{2}(\mu(1-\nu) + (1-\mu))\left(\frac{1}{2}-\phi\right)^m, \\
E[u_F] &= p_x XF + (1-p_x)(\mu\nu\cdot 0 + \mu(1-\nu)YF + (1-\mu)YF) \\
&= \frac{1}{2}\left(\frac{1}{2}+\phi\right)^m + \frac{1}{2}(\mu(1-\nu) + (1-\mu))\left(\frac{1}{2}-\phi\right)^m.
\end{aligned}
$$

The imitation is favored to an unfair allocation in case of

$$E[u_{OI}] - E[u_O] = \frac{1}{2}\left(\frac{1}{2}+\phi\right)^m - \frac{1}{2}\mu(1-\epsilon)^m \overset{!}{\geq} 0$$

Equalizing with zero and solving to μ yields

$$\Rightarrow \quad \mu^* = \frac{\left(\frac{1}{2}+\phi\right)^m}{(1-\epsilon)^m}.$$

The share μ^* is the borderline between a positive and a negative sign. For all $\mu < \mu^*$ the share of fair agents is high enough to induce an incentive to select the imitating opportunistic rule OI instead of O. Comparing the unfair opportunistic rule with fair behavior we have an incentive to play fair if

$$E[u_F] - E[u_O] = \frac{1}{2}\left(\frac{1}{2}+\phi\right)^m - \frac{1}{2}\mu((1-\epsilon)^m - \nu\epsilon^m) \overset{!}{\geq} 0$$
$$\Rightarrow \quad \mu^{**} = \frac{(\frac{1}{2}+\phi)^m}{(1-\epsilon)^m + \nu\epsilon^m} \leq \mu^*.$$

Again, μ^{**} is the borderline between a positive and a negative sign. This means that for all $\mu < \mu^{**}$ the fair rule outperforms the unfair (non-imitating) opportunistic rule. Because of $\mu^{**} \leq \mu^*$ this condition is more restrictive than the condition for imitation. Furthermore, μ^* and μ^{**} depend on the degree of fairness α (see figure 1).

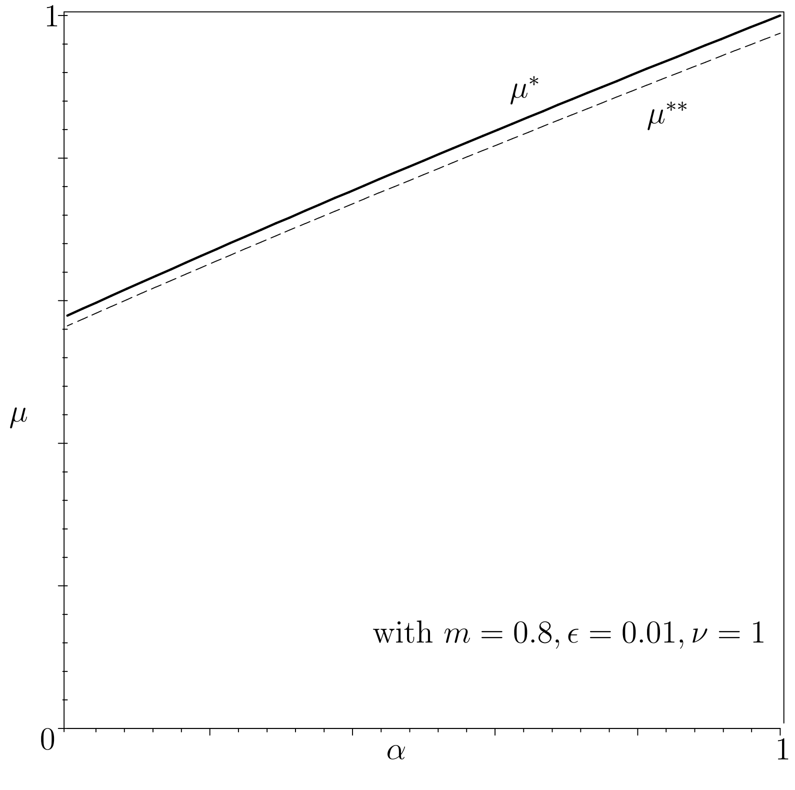

Figure 1: Critical values μ^* and μ^{**} in case of anonymous interactions

At last we have to compare the imitating opportunistic rule with fair behavior. Obviously,

$$E[u_F] - E[u_{OI}] = -\frac{1}{2}\mu\nu\epsilon^m \leq 0$$

holds true. This means that for $\nu > 0$ there is never an incentive to change from opportunistic imitation to the fair rule and there is indifference in case of $\nu = 0$.
Consider $\mu < \mu^*$. Then *all* opportunistic players have an incentive to imitate, that means they adopt rule OI. Hence, it is $\nu = 0$ for all $\mu < \mu^*$ and, conversely, $\nu = 1$ for all $\mu > \mu^*$. In case of $\nu = 0$, however, we have $\mu^{**} = \mu^*$ which implies that there is indifference between OI and F. Therefore we obtain two types of behavioral equilibrium profiles:

- The profile $(O, ..., O)$ or resp. $\mu = 1$ where all allocations are unfair,
- A continuum of profiles containing OI and F with $\mu < \mu^* = \mu^{**}$ (we neglect the special case of indifference $\mu = \mu^*$).

In case of complete fairness ($\alpha = 0$) we have

$$\mu^*_{max} = \left(\frac{1}{2(1-\epsilon)}\right)^m \simeq \left(\frac{1}{2}\right)^m$$

which is the maximum share of opportunistic (imitating) players in a mixed behavioral equilibrium profile. Since in a mixed equilibrium profile all opportunistic players imitate we will never observe unfair allocations. This is unrealistic. The main difference to the case of complete information (local group) is that the advantage of playing fair or to imitate is not (only) due to risk aversion.

2.4 Interactions with partially observable rules

No real individual interacts the whole life exclusively with anonymous agents or exclusively within a well-known local group. If an internalized social norm or a manifest attitude should make sense it has to perform well in a world where sometimes anonymous interactions and somtimes interactions in a local group occur. Depending on personal and social circumstances the frequency of local and anonymous interactions differs from agent to agent. Let q_i ($0 < q_i < 1$) be the probability for player i to interact with a member of a local group, and hence, $(1 - q_i)$ is the probability that the randomly chosen opponent is anonymous. We calculate the expected utilities of all rules in the same way as above:

$$\begin{aligned}
E[u_O] &= q_i \left[p_x(\mu XO + (1-\mu)XF) + (1-p_x)(\mu YO + (1-\mu)YF)\right] + \\
&\quad (1-q_i)\left[p_x(\mu XO + (1-\mu)\cdot 0) + (1-p_x)(\mu(\nu YO + (1-\nu)YF) + (1-\mu)YF)\right] \\
&= q_i \left[\frac{1}{2}\mu((1-\epsilon)^m + \epsilon^m) + \frac{1}{2}(1-\mu)\left(\left(\frac{1}{2}+\phi\right)^m + \left(\frac{1}{2}-\phi\right)^m\right)\right] + \\
&\quad (1-q_i)\left[\frac{1}{2}\mu(1-\epsilon)^m + \frac{1}{2}\mu\left(\nu\epsilon^m + (1-\nu)\left(\frac{1}{2}-\phi\right)^m\right) + \frac{1}{2}(1-\mu)\left(\frac{1}{2}-\phi\right)^m\right], \\
E[u_{(OI}] &= q_i \left[p_x(\mu XO + (1-\mu)XF) + (1-p_x)(\mu YO + (1-\mu)YF)\right] + \\
&\quad (1-q_i)\left[p_x XF + (1-p_x)(\mu(\nu YO + (1-\nu)YF) + (1-\mu)YF)\right] \\
&= q_i \left[\frac{1}{2}\mu(1-\epsilon)^m + \frac{1}{2}(1-\mu)\left(\frac{1}{2}+\phi\right)^m + \frac{1}{2}\mu\epsilon^m + \frac{1}{2}(1-\mu)\left(\frac{1}{2}-\phi\right)^m\right] + \\
&\quad (1-q_i)\left[\frac{1}{2}\left(\frac{1}{2}+\phi\right)^m + \frac{1}{2}\mu\left(\nu\epsilon^m + (1-\nu)\left(\frac{1}{2}-\phi\right)^m\right) + \frac{1}{2}(1-\mu)\left(\frac{1}{2}-\phi\right)^m\right], \\
E[u_F] &= q_i \left[p_x XF + (1-p_x)YF\right] + \\
&\quad (1-q_i)\left[p_x XF + (1-p_x)(\mu(\nu\cdot 0 + (1-\nu)YF) + (1-\mu)YF)\right] \\
&= q_i \left[\frac{1}{2}\left(\frac{1}{2}+\phi\right)^m + \frac{1}{2}(\frac{1}{2}-\phi)^m\right] + \\
&\quad (1-q_i)\left[\frac{1}{2}\left(\frac{1}{2}+\phi\right)^m + \frac{1}{2}\mu(1-\nu)\left(\frac{1}{2}-\phi\right)^m + \frac{1}{2}(1-\mu)\left(\frac{1}{2}-\phi\right)^m\right].
\end{aligned}$$

Since q_i is a specific parameter for each player, the incentives to imitate or to play fair are individually different. For an opportunistic player there is an incentive to imitate if

$$\begin{aligned}
E[u_{OI}] - E[u_O] &= \frac{1}{2}(1-q_i)\left(\frac{1}{2}+\phi\right)^m - \frac{1}{2}\mu(1-q_i)(1-\epsilon)^m \overset{!}{\geq} 0 \\
\Rightarrow \quad \mu^* &= \frac{\left(\frac{1}{2}+\phi\right)^m}{(1-\epsilon)^m}.
\end{aligned}$$

Again, μ^* is the borderline between a positive and a negative sign of the expression. The incentive to adopt F instead of O is given in case of

$$\begin{aligned}
E[u_F] - E[u_O] = \frac{1}{2}(1-(1-mu)q_i)\left(\frac{1}{2}+\phi\right)^m - \frac{1}{2}\mu\epsilon^m(q_i + (1-q_i)\nu) + \\
\frac{1}{2}q_i\mu\left(\frac{1}{2}-\phi\right)^m - \frac{1}{2}\mu(1-\epsilon)^m \overset{!}{\geq} 0
\end{aligned}$$

$$\Rightarrow \quad \mu_i^{**} = \frac{\left(\frac{1}{2}+\phi\right)^m(1-q_i)}{-q_i\left(\frac{1}{2}+\phi\right)^m + q_i\epsilon^m - q_i\left(\frac{1}{2}-\phi\right)^m + (1-\epsilon)^m + (1-q_i)\nu\epsilon^m}.$$

Because the critical value μ^{**} depends on the individual parameter q_i we write $\mu^{**}(q_i)$. Finally, the incentive to change from opportunistic imitation to rule F is given if

$$E[u(F)] - E[u(OI)] =$$
$$-\frac{1}{2}q_i\mu((1-\epsilon)^m + (1-\nu)\epsilon^m) + \frac{1}{2}q_i\mu\left(\frac{1}{2}+\phi\right)^m - \frac{1}{2}q_i\mu\left(\frac{1}{2}-\phi\right)^m - \frac{1}{2}\mu\nu\epsilon^m \overset{!}{\geq} 0.$$

The sign of this expression is independent from μ. Equalizing with zero and solving to q_i yields

$$\Rightarrow \quad q_i^* = \frac{\nu\epsilon^m}{-(1-\epsilon)^m + \left(\frac{1}{2}+\phi\right)^m - (1-\nu)\epsilon^m + \left(\frac{1}{2}-\phi\right)^m} \geq 0.$$

In case of $\nu = 0$ and $q_i > q_i^* = 0$ all players have an incentive to adopt the rule F. Therefore, with $\mu < \mu^*$ the share of unfair opportunistic players is $\nu = 0$ and there is an incentive to adopt rule F. This implies $\mu = 0$ and a monomorphic population of fair players is a behavioral equilibrium profile.

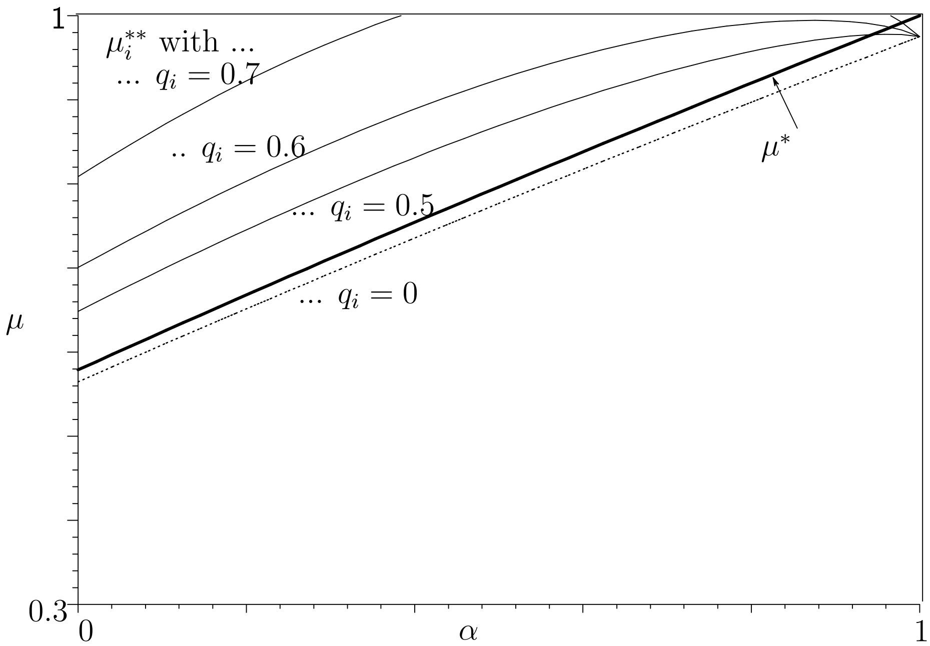

Figure 2: Critical values μ^* and $\mu^{**}(q_i)$ for mixed interactions

If $\mu > \mu^*$ there is no incentive for an opportunist to imitate and hence $\nu = 1$. It depends on q_i whether it is $\mu < \mu^{**}(q_i)$ or $\mu > \mu^{**}(q_i)$. Figure 2 depitcs μ^* and $\mu^{**}(q_i)$ with alternative values for q_i. For very small values of q_i it is $\mu^{**}(q_i) < \mu^*$ for all α. This is the case of (almost) complete anonymous interaction. Obviously for all μ with $\mu^{**}(q_i) < \mu < \mu^*$ there is $\nu = 0$ so that all $\mu^{**}(q_i)$-graphs which are paratemrized with $\nu = 1$ are relevant only in the region above μ^*. In this parameter region we have two cases:

- If $\mu^* < \mu < \mu^{**}(q_i)$ then rule O outperforms OI but rule F outperforms O. Hence, player i would adopt F.
- If $\mu^* < \mu^{**}(q_i) < \mu$ then rule O outperformas OI and F. Player i would adopt the opportunistic unfair rule O.

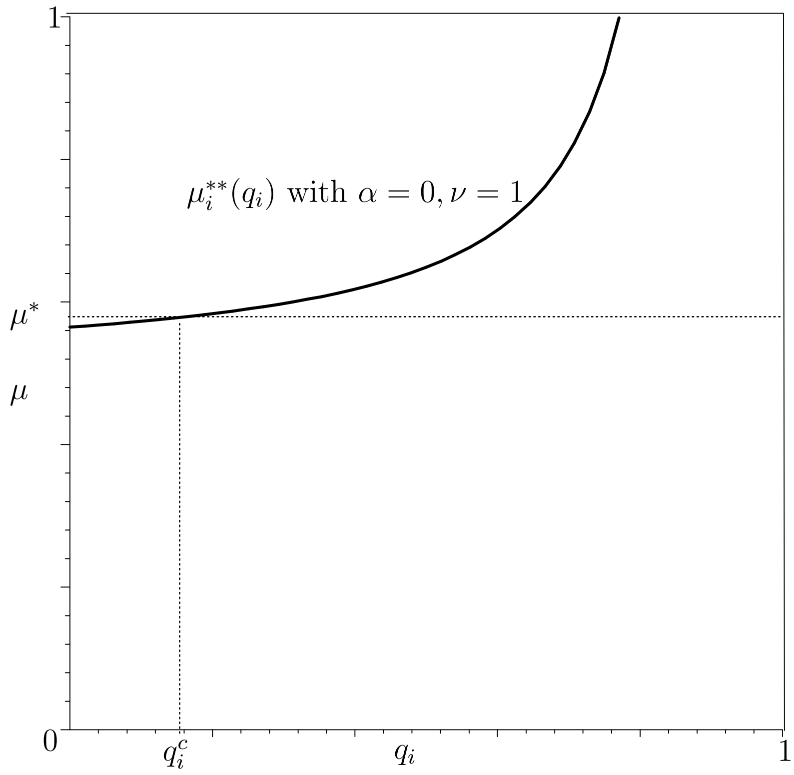

Figure 3: The border between choosing F and O

In contrast to the case of purely anonymous interactions we have for all fairness degrees α a set of mixed behavioral equilibrium profiles containing O and F with $\mu > \mu^*$ and with $\mu > \mu^{**}(q_i)$ for all opportunistic players i and $\mu < \mu^{**}(q_j)$ for all fair player. With a given α there is a nonlinear relationship between q_i and $\mu^{**}(q_i)$. For $\alpha = 0$ and $\nu = 1$ the graph is depicted in figure 3. The branch in the range $[0, q_i^c]$ has no meaning since for these (μ, q_i)-combinations it is $\nu = 0$. For all $\mu > \mu^*$ it can be seen that the higher the frequency of interaction in a local group the more likely it is to be on the right side of the graph, i.e. to adopt the fair rule F. For socially less integrated individuals who are not often interacting with well-known "trustworthy" people (q_i is low) it is more likely to adopt the unfair opportunistic rule O. In mixed behavioral equilibrium profiles in the model with local and anonymous interactions the *minimum* share of unfair players in case of $\alpha = 0$ is again $\mu^*_{min} \simeq (\frac{1}{2})^m$.

3 Discussion

Using the concept of behavioral equilibrium profiles which accounts for rational strategic behavior as well as for heuristic or norm-guided behavior we have analysed the ultimatum game with two possible behavioral rules. Depending on the information status regarding the rules which reflects the (absence of) familiarity between players we found the following results:

- In case of observable rules (interaction within a local group) there is a unique monomorphic behavioral equilibrium profile with $\mu = 0$ (only F). This is not compatible with the experimental results.

- In case of purely anonymous interactions we have one monomorphic equilibrium profile with $\mu = 1$ (only O) and a continuum of equilibrium profiles containing OI and F with $\mu < \mu^*$. Since there are no behavioral differences to observe in such a profile (only fair allocations), this is an unrealistic case.

- If the adoption of behavioral rules depends on interactions within a local group *and* with anonymous agents, the analysis leads to complete different results. There exists one monomorphic behavioral equiliobrium profile with $\mu = 0$ (only F) and a continuum of equilibrium profiles containing F and O with $\mu > \mu^*$, depending on the distribution of the q_i.

It has to be remarked that the assumed rules are rather simple. It is possible to introduce other fairness rules (e.g. where the degrees of proposed and accepted deviations from equal split differ) or to enlarge the set Ω of possible rules. Furthermore it will be interesting to model an endogenous determination of the degree of fairness α. This may shed some light on the problem how equity norms emerge in a purely opportunistic population (if this is seen as a plausible "primitive state" of a population). Starting with a low degree of fairness ($\alpha \to 1$) the critical values μ^* and μ^{**} are very high. Therefore a very small subpopulation of fair agents can invade and trigger the evolution of higher degrees of fairness and the increasing adoption of rule F. The analysis may also be applied to variants of the game like the three-person-ultimatum game (cf. Güth/van Damme 1998) or the dictatorship-game.

The crucial point of the paper is the explanation why agents adopt certain behavioral rules. It was argued that each profile of rules determines a (in this case: unique) behavioral equilibrium, and that a profile itself constitutes an equilibrium when no player can benefit from adopting another rule. It does not play a role whether the adoption comes from strategic deliberation or from adaptive processes. As we talk about "incentives" to adopt a certain rule, it might be irritating at the first sight that a behavioral equilibrium profile implies some kind of maximization calculus over the set of rules Ω. Lipman (1991) addresses the question whether this kind of recursion runs into logical problems, and he denies the question. Moreover, Rubinstein (1998) claims that there is a need for theories of boundedly rational behavior which consider that agents reason about *how* they decide. The calcules of selecting the best performing rule requires a closed and well-defined set Ω. It has to be underlined that for analytical reasons this set of rules is taken as exogeneously given. In fact, behavioral rules are not *given data* for the agents, but they are *created* by them. However, incorporating endogeneous creation of new patterns would be a non-accomplishable task of a formal theory.

It is obvious that rational maximizing behavior turns out *not* to be the (unique) best performing rule as it was often claimed to justify neo-classical assumptions as the result of evolutionary selection (cf. Alchian 1950, Friedman 1953). The concept bridges the gap between rational choice and behavioral explanations in economics. On the one hand, rational maximizing behavior as well as behavioral approaches of decision making are explained (not presumed) within an integrated framework. On the other hand, psychological and other aspects are incorporated in well-defined behavioral rules and the analysis of equilibrium profiles is a rigorous (utility based) concept which leads to clear formal propositions – this may be a charming invitation for neo-classical theory to abandon its apriorism.

How is this approach related to other explanations of fairness, especially to the indirect evolutionary approach? First of all, there are no cosiderations about special preference structures and utility functions. It is claimed that sophistication of utility functions might be a possible but not always promising way. Since the explanandum is completely unobservable one can reply to the objection of an ad hoc assumption only if preferences itself are explained like in the indirect evolutionary approach. As it has been discussed above, the latter exhibits the problem that is is doubtful to separate individual valuation and the objective outcome which drives the selection process in the proposed way. It was argued that an individual adaption or learning process can directly address the behavioral rules instead of the preferences. This has also the advantage that *heuristic* processes of reasoning and decision making are permitted like it is suggested in the bounded rationality literature. For each appointed context the behavioral rules require a detailed hypothesis which is conform to the observed results. In this paper, instead, we have taken just a very simple case of a fairness rule in order to demonstrate how the behavioral equilibrium concept can be applied to the ultimatum game.

References

Alchian, A. (1950), Uncertainty, Evolution, and Economic Theory. *Journal of Political Economy* 58, pp. 211.

Becker, G.S. (1976), Altruism, egoism, and genetic fitness: Economics and sociobiology. *Journal of Economic Literature* 14, 817-826.

Becker, G.S., Stigler, G. (1977), De Gustibus Non Est Disputandum. *American Economic Review* 1(?), 76-90.

Bolton, G.E., Ockenfels, A. (2000), ERC - A Theory of Equity, Reciprocity and Competition. *American Economic Review* 90(1), 166-193.

Camerer, C., Thaler, R. (1995), Ultimatums, Dictators, and Manners. *Journal of Economic Perspectives* 9, 209-219.

Falk, A., Fischbacher, U. (1998), A Theory of Reciprocity. mimeo.

Forsythe, R., Horowitz, J.L., Savin, N.E., Sefton, M. (1994), Fairness in Simple Bargaining Games. *Games and Economic Behavior* 6, 347-369.

Friedman, M. (1953), *Essays in Positive Economics.* Chicago: Chicago University Press.

Güth, W. (1995), On ultimatum bargaining experiments - A personal review. *Journal of Economic Behavior and Organization* 27, 329-344.

Güth, W., Kliemt, H. (1998), The indirect evolutionary approach: Bridging the gap between rationality and adaptation. *Rationality and Society* 10(3), 377-399.

Güth, W., Pull, K. (2002), Will Equity Evolve? An Indirect Evolutionary Approach. mimeo.

Güth, W., van Damme, E. (1998), Information, Strategic Behavior, and Fairness in Ultimatum Bargaining: An Experimental Study. *Journal of Mathematical Psychology* 42, 227-247.

Huck, S. (1997), Ultimatum bargaining and beyond – From behavioral regularities to evolutionarily stable preferences. Aachen: Shaker.

Lipman, B. (1991), How to decide how to decide how to...: modelling limited rationality. *Econometrica* 59, 1105-1125.

Pasche, M. (2001), Equilibrium Concepts for Boundedly Rational Behavior in Games. Working Paper No. 2001/03, University of Jena, Economics Faculty.

Rabin, M. (1993), Incorporating Fairness into Game Theory and Economics. *American Economic Review* 83, 1281-1302.

Rubinstein, A. (1998), *Modelling Bounded Rationality.* Cambridge, Mass.: MIT-Press.