ELSEVIER

Journal of Economic Behavior & Organization
Vol. 61 (2006) 617–631

JOURNAL OF
Economic Behavior
& Organization

www.elsevier.com/locate/econbase

On the behavior of proposers in ultimatum games

Thomas Brenner [a], Nicolaas J. Vriend [b,*]

[a] *Max Planck Institute, Jena, Germany*
[b] *Queen Mary, University of London, Department of Economics, Mile End Road, London E1 4NS, UK*

Received 28 September 2003; accepted 21 July 2004
Available online 12 September 2006

Abstract

We demonstrate that one should *not* expect convergence of the proposals to the subgame perfect Nash equilibrium offer in standard ultimatum games. First, imposing strict experimental control of the behavior of the receiving players and focusing on the behavior of the proposers, we show experimentally that proposers do not learn to make the expected-payoff-maximizing offer. Second, we put the experimental data into perspective by considering a range of learning theories (from approximately optimal to boundedly rational). These theoretical benchmarks explain that the lack of convergence to the minimal offer is an inherent feature of the learning task faced by the proposers.

JEL classification: C72; C91; D81; D83

Keywords: Ultimatum game; Non-equilibrium behavior; Laboratory experiment; Multi-armed bandit; Optimal learning; Gittins index; Bounded rationality

1. Introduction

One of the games most extensively studied in the literature in recent years is the ultimatum game. The reason that this game is so intriguing seems to be that the game-theoretic analysis is straightforward and simple, while the overwhelming experimental evidence is equally straightforward but at odds with the game-theoretic analysis (see, e.g., Güth et al., 1982; Güth and Tietz, 1990; Thaler, 1988).

In the basic ultimatum game there are two players and a pie. Player A proposes how to split the pie between herself and player B. Upon receiving player A's proposal, player B has two options:

* Corresponding author. Tel.: +44 20 7882 5081; fax: +44 20 8983 3580.
E-mail address: n.vriend@qmul.ac.uk (N.J. Vriend).

0167-2681/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jebo.2004.07.014

to accept the proposal, which will then be carried out, or to reject it, after which both get nothing. Many variants of this basic setup have been considered in the literature. There are many Nash equilibria in this game. Every strategy for player A combined with any strategy for player B that accepts that offer but rejects all lower offers is one. But there is a unique subgame perfect equilibrium: player A offers the minimal piece, and player B accepts that.[1]

Empirical evidence shows time and again that this is not what happens in the laboratory. Players A usually offer somewhat less than half the pie to players B, and players B usually reject small offers. Concerning player A's behavior, there are two main explanations for this anomaly offered in the literature. First, some argued that fairness and reciprocity considerations are the forces driving players A to offer more than the standard game-theoretic analysis would suggest (see, e.g., Forsythe et al., 1994). An alternative explanation found in the literature is that players A are basically following an adaptive, best-reply seeking approach to the behavior of players B. In a multi-period setup where players played the game repeatedly but each time against different players, some papers showed how it can happen that players A unlearn to play the subgame perfect equilibrium strategy as players B have not learned yet that they should play their perfect equilibrium strategy. Once players A do not play that strategy anymore, players B will never learn to play theirs. Such learning dynamics are shown in Roth and Erev (1995), who follow a reinforcement learning approach, and Gale et al. (1995) who use replicator dynamics.

Both explanations are somehow based on the assumption that players A learn to play best-replies to the behavior of players B. The deviation from the prediction of subgame perfect Nash equilibrium is, therefore, mainly explained by the deviating behavior of players B. Players A only react to this deviation, which is caused either by a slow learning process or by fairness considerations.[2] In this paper, in contrast, we will show that the adaptive behavior of players A may also cause deviations from the subgame perfect Nash equilibrium, independent from the adaptive behavior of players B.[3] That is, we show that the learning task faced by players A is such that one should expect players A to stay away from the optimal minimal offer even when the behavior of players B qualitatively resembles perfect equilibrium behavior.

In order to focus on the behavior of players A, we design an ultimatum game experiment in which the behavior of a large population of players B is fixed by some computer algorithm. This was known to the players. Our experimental design has two advantages. First, as there are no payoffs to other people influenced by the behavior of players A, fairness considerations cannot play a role. Second, learning in ultimatum games is essentially a coevolutionary process: players learn about the behavior of other players who learn about the behavior of other players who learn and so on. Our experimental design allows us to focus on the learning behavior of players A, abstracting from the complications and peculiarities related to coevolutionary processes.[4]

Basically, the problem faced by a player A is a multi-armed bandit problem. There are four treatments that differ in the general level of acceptance rates. The experimental parameters in each

[1] Strictly speaking, in case it is a discrete choice problem including zero, there are two subgame perfect equilibria, with player A offering either zero or the smallest possible strictly positive piece to player B.

[2] A recent example of the former is in Cooper et al. (2003).

[3] Vriend (1997) presented some theoretical considerations why paying more attention to the learning behavior of the proposing players as such could be worthwhile.

[4] Our approach is similar in spirit to a dictator game (see, e.g., Bolton et al., 1998), in the sense that dictator games were also invented to cut out players B. But there are two advantages of our setup. First, in a dictator game players A know the behavior of players B (accepting anything), whereas this is not the case in our setup. Second, in a dictator game fairness considerations may still play a role.

treatment are such that two monotonicity properties are satisfied: higher offers are more likely to be accepted while lower offers give higher expected payoffs to the proposer.

Two stylized facts stand out in the experimental data. First, although the experiment comprises 100 periods, there is only a small tendency for the average offer to come down to the minimal offer, the one that maximizes a proposer's expected payoff. Second, although the incentive structure of the proposers is the same in each treatment, lower general rejection rates lead to significantly lower offers.

We consider a range of learning theories from approximately optimal learning (based on the Gittins index) to more boundedly rational learning methods. These learning theories provide us with theoretical benchmarks, putting the experimental data into perspective, as they give us an idea about what we could reasonably expect learning behavior to lead to in the situation faced by the players in the experiment. As it turns out, both stylized facts would be predicted by these learning theories.

In other words, even if the behavior of the receiving players qualitatively resembles subgame perfect equilibrium behavior (such that the incentive structure for the proposing players is the same), one should *not* expect convergence to the subgame perfect Nash equilibrium in standard ultimatum games. This offers an explanation for the lack of convergence to the subgame perfect Nash equilibrium in ultimatum games, an explanation that complements most existing explanations.

The rest of this paper is organized as follows. In Section 2 we present the experimental design, and in Section 3 the experimental results. Various learning theories to explain the experimental data are discussed in Section 4, and their predicted dynamics are examined in Section 5. Section 6 concludes.

2. Experimental design

The underlying idea for our design is the following. First, we wanted to set up a stylized ultimatum game in which the optimal strategy for players A would coincide with the subgame perfect equilibrium strategy of the standard ultimatum game of offering only a minimal slice. Second, as we wanted to focus on players A's learning behavior, we wanted to be in a position to exclude as much as possible other well-known explanations for players A staying away from the optimal action. In particular, this implies that we needed to be in a position to abstract from the learning behavior of the receiving players B.

We play an ultimatum game in which the pie has size 9, and we allow only integers from 1 to 9 to be chosen as offers (see also Roth and Erev). The experimental subjects are players A, who play against players B who form a large population of artificial agents and make their decisions using some computer algorithm. Every period a given player A is randomly matched with a player B that he has not met yet. Player A enters his offer, and then the reply of player B and the corresponding payoff for player A are shown. There are 100 periods to be played. Notice that this is more periods than in experimental ultimatum games typically reported in the literature. Fig. 1 shows a player's screen during the experiment. A player could at any moment during the experiment scroll through his complete history. The identity of players B is listed on the screen to make clear that in every period the opponent is a different player. Each period, after the choice of player A, it takes 5–15 s (uniform randomly chosen) before the reply of player B is listed (saying "please wait for reply player B"). This suggests players B make serious choices, and it avoids players A getting rushed too much by the speed of players B.

Fig. 1. Sample interface during experiment.

Table 1
Acceptance probabilities

Treatment	Offers								
	1	2	3	4	5	6	7	8	9
ult3	0.30	0.31	0.32	0.35	0.39	0.47	0.64	1.00	1.00
ult5	0.50	0.51	0.54	0.58	0.66	0.79	1.00	1.00	1.00
ult7	0.70	0.72	0.76	0.82	0.92	1.00	1.00	1.00	1.00
ult10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Using artificial players B allows us to control the environment for players A. Players A are told that players B are artificial players. Given that the players B are artificial players, altruism considerations are irrelevant. We told players A: *"Each of those players B's behavior is systematic in the following sense: If a specific player B has accepted an offer x then that player B would have accepted as well any offer greater than x. And if offer x had been rejected by that player B, so would have been all offers smaller than x. Of course, different players B might have different opinions about which offers are acceptable or not. The players B do not change their behavior over time"* (see instructions in the Appendix in Supplementary Material, available on the JEBO website). As a result, the population of players B can be characterized by a probability density function that a given offer will be accepted. The probability that a given offer is accepted is monotonically increasing in the size of the offer.[5] We organize four treatments, which differ in the general level of acceptance probabilities. The probabilities that a randomly chosen player B will accept a given offer in each treatment is listed in Table 1.

These probabilities are based on the following considerations. First, the expected payoff maximizing offer is 1, which coincides with the subgame perfect equilibrium strategy of a standard

[5] Hence, in our experiment there is a heterogeneous population of players B who play pure strategies characterized by a reservation value property. Every player B has a reservation value, but different players may have different values. Given the size of the population (one million players), the following alternative interpretation is approximately correct. The population of players B consists of identical players using a mixed strategy characterized by acceptance probabilities that are monotonically increasing in the size of the offer.

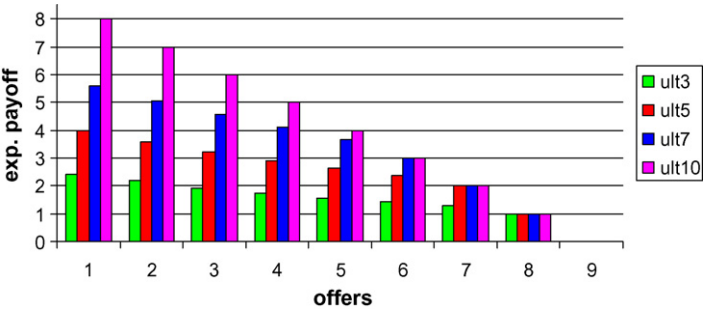

Fig. 2. Expected payoffs across treatments.

ultimatum game.[6] Second, the acceptance probabilities increase monotonically with the size of the offer, thus maintaining realistic assumptions concerning the behavior of players B. Third, given the two considerations already mentioned, we wanted to make the learning task as easy as possible. Therefore, the minimal offer of 1 gives an expected payoff that is clearly higher than for any other possible offer. Moreover, we avoid the existence of sub-optimal peaks in the range of possible offers, which would make it more difficult for players to reach the global optimum. Starting from the minimal offer of 1, in each treatment each next higher offer gives an expected payoff that is at least 10 percent lower. Fig. 2 shows these resulting expected payoffs for each treatment. Notice that the ult10 treatment corresponds exactly to the subgame perfect equilibrium behavior of players B, and that the payoff structure in each of the other treatments is very similar.

Most of the experiment (treatments ult3, ult5, and ult7) was conducted in the computerized experimental laboratory CEEL of the University of Trento in November 1997, with one additional treatment (ult10) organized at Queen Mary, University of London in March 2004.[7] With only a few exceptions, all players for a given treatment were simultaneously in the laboratory. The players, 19 for the ult3 treatment, 20 for the ult5 and ult7 treatments, and 14 for the ult10 treatment, went through the experiment in about an hour. The exchange rates were 83.3 Italian Lires per point for ult3, 50.0 Lit./point for ult5, 35.7 Lit./point for ult7, and 1 pence/point for the ult10 treatment.[8] These exchange rates were chosen such that the monetary incentives were essentially the same in each treatment. The differences are due to rounding and to the fact that the acceptance probabilities cannot exceed 1, which is of relevance mainly for some of the highest offers.

3. Experimental data

Fig. 3 presents the time-series of the offers averaged over the players in each of the four treatments.

[6] Whether the expected-payoff-maximizing offer is also the optimal offer (given the acceptance probabilities) may depend on a player's risk-attitude. Following the arguments presented in Rabin (2000), within the framework of Expected Utility theory people cannot be risk-averse for such small stakes as offered in our experiment because it would imply absurd risk-aversion for large stakes. Hence, risk might become an issue only if we were to abandon Expected Utility theory, which is an issue beyond the scope of this paper.

[7] The additional treatment was conducted as a control experiment. Since it was conducted at a different place at a different time, its results cannot be compared directly to those of the other experiments. The main aim of this additional treatment was to check whether players converge to the optimal choice in a situation in which all uncertainty was removed. There seems little reason to believe that the answer to this question depends on the specific place and time.

[8] In addition, players in the ult10 treatment received a £5.00 show-up fee.

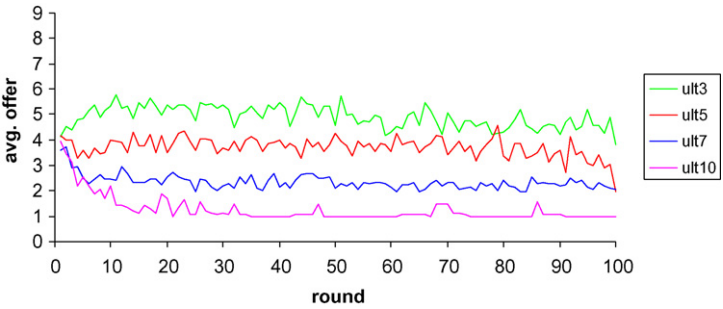

Fig. 3. Average offers in ultimatum game experiment.

We observe the following. First, in each treatment the average initial offer is about 4, slightly below the 50 percent offer. Second, although the monetary incentives were the same in each treatment, there is a systematic difference across treatments, with lower rejection probabilities leading to lower offers.[9] These differences emerge in particular during the first 10 periods, with the average offers in the ult10 treatment reaching 1 after 20 periods. Third, from period 10 to 90, the average offers in all treatments decline significantly but slowly,[10] and in none but the ult10 treatment do the players learn to make the minimal offer of 1 (although there is a downward end effect, in particular in the treatments ult3 and ult5).[11]

4. Some learning theories

In this section we present a number of learning theories that might help to put the experimental observations into perspective. That is, these learning theories serve as benchmarks to give us an idea about what learning might reasonably be expected from the players.

4.1. Modeling rational learning

As explained above, in our experimental design we simplified the standard ultimatum game into a one-person decision problem. This decision problem in our experiment resembles a multi-armed bandit. For many specific kinds of repeated multi-armed bandit situations optimal behaviors are given in the literature (see, e.g., Gittins, 1989; Bergemann and Välimäki, 2001; Brezzi and Lai, 2002). One standard multi-armed bandit situation is characterized by people knowing the value of the payoffs, π_i, that they might receive with each of the arms i. They also know in this situation that they receive each payoff with a fixed but unknown probability p_i. If, in addition,

[9] A robust rank-order test, based on the average offer for each individual player over the 100 periods of the experiment, shows that the players offer significantly more in the ult3 than the players in the ult5 treatment, who in turn offer significantly more than those in the ult7 treatment, with the latter offering significantly more than those in the ult10 treatment (all at 0.000005 significance level; 1-sided).

[10] A simple regression against time gives coefficients of −0.012, −0.004, −0.004 and −0.003 for the ult3, ult5, ult7 and ult10 treatments, respectively (all significant at 0.005; 1-sided). Extrapolation of the observed rates implies that it would take hundreds of periods more in the ult3, ult5 and ult7 treatments for the average offer to reach 1.

[11] Such unexplained end effects (not related to players who stop reputation building activity if few or no further interactions are expected) are relatively common in experiments and might be a topic for further investigation.

the probabilities p_i are independent of each other, the *Gittins index* can be used to determine the optimal behavior.

The *Gittins index*, introduced by Gittins (1979, 1989) assigns at each time to each arm the maximal average payoff that can be obtained by repeatedly choosing that arm for an event-dependent number of times. Each time an arm is chosen, it is determined randomly, according to the probability assigned to the arm, whether a payoff is obtained or not. After each choice it is decided, depending on the experience with this arm, whether the arm is chosen again or not. How this decision is made is called the *stopping rule*. Many different stopping rules can be imagined. If we consider the time span from the actual time t until the time at which the choice is changed because of the stopping rule, the average payoff that is received within this time span can be calculated. This average payoff depends on the stopping rule. The Gittins index is defined as the maximal average payoff that can be reached by any stopping rule. We denote the Gittins index for arm i at time t by $g_i(t)$. To calculate the average payoffs for each stopping rule it is necessary to calculate the probability for obtaining a payoff at each time. Since the real probabilities are not known, Bayesian updating is used for calculating these expected probabilities. Hence, the concept of Gittins indices is based on two basic features. First, Bayesian updating is used to calculate the expected probabilities of payoffs (the initial belief is that all relevant hypotheses are equally likely). Second, average expected payoffs are calculated for each arm separately according to the stopping rule approach. Gittins has proved that choosing at each time, t, the arm, i, with the highest Gittins index, $g_i(t)$, is the optimal strategy for the situation described above.

Strictly speaking, the use of the Gittins index is not appropriate for the situation faced by the players in our experiment. The reason is that the arms are not independent. The information given to the players implied that the probability of acceptance was weakly increasing in the size of the offer. This requires two modifications of the standard Gittins index approach.

First, when updating the probability of acceptance for some arm i in the context of our ultimatum game, a player should also update the probabilities for all other arms, as he knows that $p_i \leq p_j$ for each $i<j$. As a consequence the hypotheses in Bayesian learning have to be formulated for all arms jointly. Hence, a hypothesis is characterised by nine probabilities: $p_1, p_2, \ldots, p_9$, the probabilities assigned to each of the nine arms. The number of possible hypothesis is reduced by the condition $p_1 \leq p_2 \leq p_3 \leq p_4 \leq p_5 \leq p_6 \leq p_7 \leq p_8 \leq p_9$. Therefore, the set of all feasible hypothesis is given by

$$H = \{(p_1, p_2, \ldots, p_9) | p_1 \leq p_2 \leq p_3 \leq p_4 \leq p_5 \leq p_6 \leq p_7 \leq p_8 \leq p_9\}.$$

The probability $P(h, t)$ that is assigned to each hypothesis, $h \in H$, at each time, t, is updated according to Bayes' rule. The expected probability to obtain a payoff if choosing arm i is given by

$$E_i(t) = \sum_{h \in H} P(h, t) \times p_i(h)$$

or

$$E_i(t) = \int_0^1 \int_0^{p_9} \int_0^{p_8} \int_0^{p_7} \int_0^{p_6} \int_0^{p_5} \int_0^{p_4} \int_0^{p_3} \int_0^{p_2} P(p_1, p_2, \ldots, p_9, t) p_i \, \mathrm{d}p_1 \mathrm{d}p_2 \mathrm{d}p_3 \mathrm{d}p_4 \mathrm{d}p_5 \mathrm{d}p_6 \mathrm{d}p_7 \mathrm{d}p_8 \mathrm{d}p_9 \quad (*)$$

The expected probabilities defined by Eq. (*) are used in the calculation of the Gittins indices.[12] Hence, the Gittins indices are based on expected probabilities that are calculated using *all* the available information.

Second, when choosing an arm, a player should not simply try the arm with the highest Gittins index because he should take into account as well that the outcome with that arm will provide useful information about the other arms (as, according to Eq. (*), the expected payoffs of other arms change when using a given arm). No optimal strategy is known in the literature taking this second point into account.[13]

Therefore, we will compute adjusted Gittins indices (taking the first modification into account), and then let players simply choose the arm with the highest Gittins index. Hence, we obtain an approximation of optimal behavior. This approximation deviates from optimal behavior only by the fact that it assumes the expectation for the average payoffs of all other arms to remain constant while repeatedly choosing one arm. Since this deviation is similar for all arms, the ranking of the Gittins indices should be little influenced by this approximation.

The Gittins indices are calculated as described in the literature (see Gittins, 1979, 1989). At each time, t, for each arm, i, the stopping rule is calculated that leads to the highest average payoff for repeatedly choosing this arm. This is done through backward induction on all possible sequences of outcomes for repeatedly choosing arm i. Whenever continuing the choice of arm i decreases the average payoff, the choice is stopped. The expected probability $E_i(t)$ of arm i to lead to a positive payoff is calculated at each time, t, according to Eq. (*). The probability $P(h, t)$ for each hypothesis, h, is updated according to Bayes' rule. This means that

$$P(p_1, p_2, \ldots, p_9, t+1) = \frac{p_1 \times P(p_1, p_2, \ldots, p_9, t)}{\sum_{h \in H} p_i \times P(p_1, p_2, \ldots, p_9, t)}$$

holds if arm i is chosen at time t and a positive outcome results and that

$$P(p_1, p_2, \ldots, p_9, t+1) = \frac{(1-p_1) \times P(p_1, p_2, \ldots, p_9, t)}{\sum_{h \in H} (1-p_i) \times P(p_1, p_2, \ldots, p_9, t)}$$

holds if arm i is chosen at time t and an outcome of zero results. The initial prior is that each hypothesis, h, is equally likely. This means that value of $P(p_1, p_2, \ldots, p_9, 0)$ is the same for all probabilities $p_1, p_2, \ldots, p_9$ that satisfy the condition $p_1 \leq p_2 \leq p_3 \leq p_4 \leq p_5 \leq p_6 \leq p_7 \leq p_8 \leq p_9$. The initial behavior is calculated according to the adjusted Gittins index on the basis of these values.

4.2. Modeling boundedly rational learning

To put the experimental data further into perspective, we now describe some models of boundedly rational learning.

With reinforcement learning we assume that players A have no understanding of game theory, the structure of the ultimatum game, or the behavior of players B. Players A are boundedly rational agents who behave adaptively to their environment. They simply try actions and are in the future more likely to choose those offers that had been more reinforced (through higher payoffs) in the

[12] This integral can be calculated in closed form for each possible history. However, for reasons of convenience we will do this numerically.

[13] To calculate optimal behavior in such a situation requires the examination of all possible sequences of actions, their potential outcomes, and the respective probabilities, which is computationally not feasible.

past. One can imagine players A playing with a multi-armed bandit, where different arms might give different payoffs, and players A do not know at the start which is the best arm to pull.

A first basic reinforcement learning model is due to Roth and Erev. At time $t = 1$ each player has an initial propensity to choose his ith arm given by some real number $q_i(1)$. We assume $q_j(1) = q(1)$ for each j, and $\Sigma q_j(1) = 10$ (following Roth and Erev). If a player plays arm i at time t and receives a payoff of z, then the propensity to choose arm i is updated by setting $q_i(t+1) = q_i(t) + z$, while for all other arms j, $q_j(t+1) = q_j(t)$. The probability that the player selects his ith arm at time t is $p_i(t) = q_i(t)/\Sigma q_j(t)$, where the sum is over all the available arms j. Thus, given the reinforcements for all offers, a player chooses (with some experimentation) his most reinforced offer. Notice that in this basic reinforcement learning model players ignore the interdependence of the arms.

A second model of reinforcement learning was introduced into the economics literature by Kirman and Vriend (1995).[14] In this model a player computes the expected payoff for each possible arm on the basis of past payoffs actually experienced, and chooses the arm with the highest expected payoff subject to some experimentation. If we denote by $r_i(t)$ the expected payoff of arm i at time t, and a player receives a payoff of z using arm i, then his expected payoff for arm i will be updated as: $r_i(t+1) = r_i(t) - c \times r_i(t) + c \times z$, where $0 < c < 1$. Given the expected payoffs when a player chooses an arm, each arm i makes a bid b equal to its strength with a small stochastic term added: $b_i(t) = r_i(t) + \varepsilon$, where ε is a $N(0, \sigma)$ error term. The arm with the highest bid in this 'stochastic auction' wins the right to be active. Following Kirman and Vriend (1995), we set the initial expected payoffs equal for each arm at a level that corresponds to the maximum possible payoff. This ensures that each arm will be tried. Normalizing all payoffs to [0, 1], we set the standard deviation σ of the error term at 0.05, while for the adjustment parameter we will consider all values $0 < c < 1$.

Players A who behave according to learning direction theory (see, e.g., Selten and Stoecker, 1986) look at the outcome of the most recent period, and reason in which direction a better offer could have been found. They, then, simply adjust their current offer into that direction. More specifically, if a player A found an offer i was rejected at time t, then at time $t+1$ he will offer $i+1$ to player B (unless offer i equaled the maximal possible offer). If, on the other hand, offer i was accepted at time t, then at time $t+1$ he will offer $i-1$ to player B (unless offer i equaled the minimal possible offer). Notice that a player learning according to learning direction theory can be seen as seeking myopically for a best-response against his latest opponent, and that implicitly he takes the interdependence of the arms into account.

5. Predicted dynamics for learning theories

We first consider the predicted choices for the model of approximately optimal learning based on the adjusted Gittins indices, adjusted to take into account the interdependence of the arms in our ultimatum game.

Fig. 4 shows the average behavior of the model of approximately optimal learning over 100 runs for the four different treatments. We make the following observations. First, the initial choice is 4. This is due to the fact that, according to the initial probabilities for the different hypotheses, the fourth arm has the highest Gittins index in the first round. Second, differences between the treatments emerge early on, with the ult3 treatment showing an increase in average offers, the ult7

[14] This was eventually published as Kirman and Vriend (2001). Sarin and Vahid (1999) analyze some theoretical properties of this model.

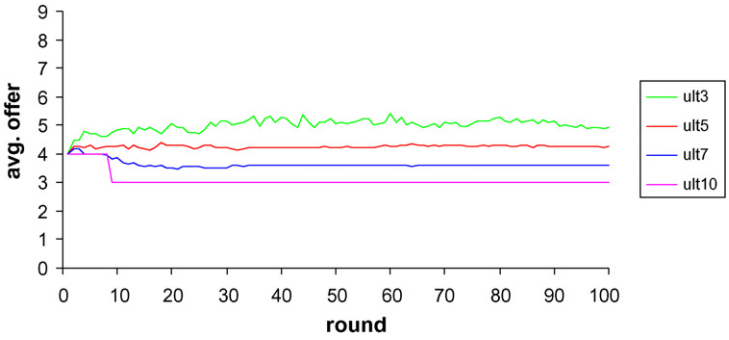

Fig. 4. Theoretical prediction of the average behavior using adjusted Gittins indices.

and ult10 treatments more of a decline, and the ult5 treatment somewhere in between. Third, after the initial learning phase there is no further downward trend and no convergence to the optimal minimal offer. The optimal choice (an offer of 1) is chosen with probability zero in each of the four treatments after 100 rounds, and the average offer falls below 4 only for the treatments ult7 and ult10.

The reason for the lack of a downward trend towards the minimal offer is a lack of experimentation of the model of approximately optimal learning. Given the structure of the ultimatum game, each outcome provides information not only about the probability behind the chosen arm but also about the probabilities behind the other arms because the probabilities depend on each other. As a result, experimentation by actually trying other arms is relatively less attractive than in the case of independent arms. In other words, the situation faced by the players in our ultimatum game experiment is such that even with approximately optimal learning, no convergence to the minimal offer of 1 takes place within 100 rounds.

Given the theoretical benchmark of the adjusted Gittins index, we now have another look at the experimental data. Notice that the predicted behavior of the model of approximately optimal learning shows qualitative similarities with the actual experimental data. This concerns in particular the initial choices, the early emergence of differences between the treatments, and the lack of convergence to the minimal offer. There is, however, also some qualitative difference between the predicted behavior of the model of approximately optimal learning and the experimental data. In the experimental data we observe a weak downward trend, whereas the model of approximately optimal learning does not show a trend at all. In the experimental data, we see that the players tend to experiment more than in the model of approximately optimal learning, but not enough to learn to choose the optimal minimal offer, with the exception of the ult10 treatment.

To gain some further insights into where we could expect learning to lead to in the ultimatum game experiment, we now consider the predicted behavior of a number of learning models that deviate in one way or another from the optimal one.

We start from the above analysis of the model of approximately optimal learning showing that players are unable to learn to make the optimal minimal offer because exploration is not sufficiently attractive due to the interdependence of the probabilities of the arms. Furthermore, the experiment shows that people experiment more than predicted by the above modeling. Hence, it would be interesting to see what theoretical benchmark we get if the arms are treated as being *independent*, neglecting the information about the relationship between the arms. The standard Gittins index approach provides such a benchmark. That is, there is a separate set of hypotheses

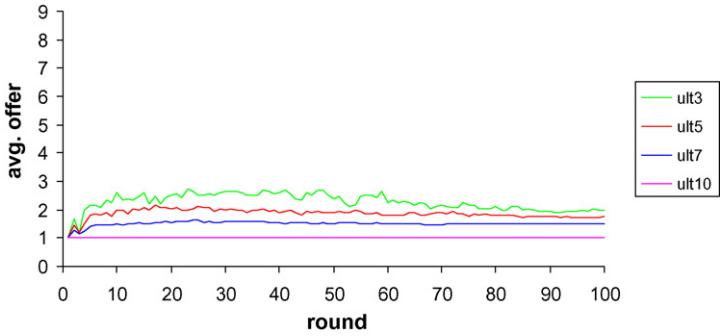

Fig. 5. Theoretical prediction of the average behavior using Gittins indices.

for each arm. The hypotheses for arm i are given by all possible probabilities p_i. The initial beliefs are that each hypothesis is equally likely. This leads to an initial prediction of 0.50 for each arm to lead to a positive outcome. The expected probability for each arm is determined only by the experience that has been made with this arm in the past. On the basis of the probabilities that result from Bayesian updating, the Gittins indices are calculated for each arm. Then, the arm that offers the highest Gittins index is chosen.

Fig. 5 shows the average behavior for 100 runs of the standard Gittins index approach, neglecting the interdependence of the arms. We make the following observations. First, the initial offer made is 1. This is due to the unbiased priors, attaching a probability of acceptance of 0.50 to each arm. Second, the early learning effect leads to increased offers in all but the ult10 treatment, with differences between the four treatments emerging. Third, this is followed by a weak downward trend in the ult3, ult5 and ult7 treatments, with only about half of the players making the optimal minimal offer in the end in these treatments (47 percent, 56 percent and 64 percent for the ult3, ult5 and ult7 treatments, respectively), although they had all started there in the first round. In the ult10 treatment, all players choose the optimal minimal offer throughout. The fact that there is more of a downward trend towards the optimal minimal offer may seem counter-intuitive, as this model makes less use of the available information than the model of approximately optimal learning analyzed above.

Hence, for this learning model that neglects the interdependence between the arms we see again that the players do not really learn to make the minimal offer, except for the noise-free ult10 treatment, and, again, there are some qualitative similarities with the actual experimental data. The latter applies in particular to the weak downward trend.

We now turn to the predicted dynamics of the boundedly rational models of learning presented in Section 4. This is not so much to test which model fits the experimental data best, but to put the experimental data even further into perspective, and to get a better idea as to where we could expect learning to lead to in the ultimatum game.

Fig. 6 shows the average offers for 1000 players using reinforcement learning as in Roth and Erev. We observe the following. First, the initial choices are about 5. Second, differences between the treatments emerge early on. Third, there is a weak downward trend, and in none of the treatments is the optimal minimal offer approached.

Fig. 7 show the results predicted for the second reinforcement learning model, as in Kirman and Vriend (1995), using the average of 1000 players. As explained in the previous section, the adjustment parameter c was a free parameter, with $0 < c < 1$. If c is large, the feedback is noisy. That

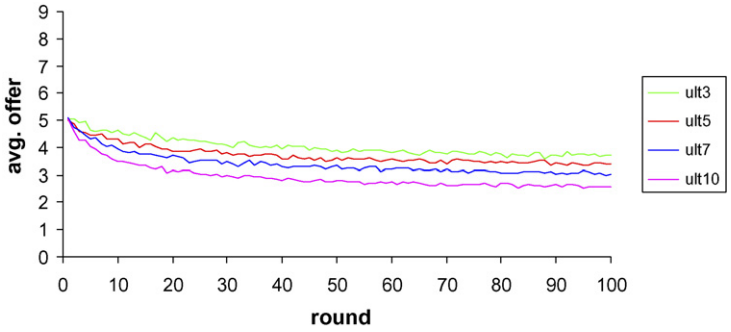

Fig. 6. Theoretical prediction of the average behavior using reinforcement learning (Roth and Erev).

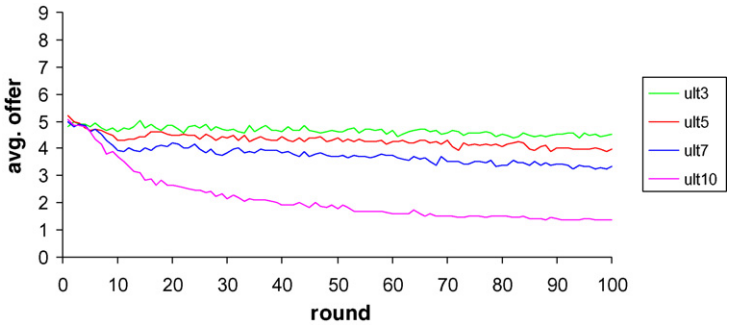

Fig. 7. Theoretical prediction of the average behavior using reinforcement learning (Kirman and Vriend, 1995).

is, as the payoffs realized are all-or-nothing, the expected payoffs are under- and overshooting all the time, leading to incorrect choices most of the time and no convergence towards the optimal arm. If c is small enough, learning is more reliable and eventual convergence to the optimal arm is assured. However, if c is too small no downward trend is observed within the 100 period window considered. What we report in Fig. 7 is the largest value for c ($c = 0.10$) such that, first, the parameter c is equal for each treatment, and second, there is a downward trend towards the optimal arm in each treatment. As we see, the model behaves very much like the other reinforcement learning model. Only the ult10 treatment is predicted to get close to the optimal minimal offer of 1, while very little convergence towards the optimal choice is predicted for the other three treatments. For other values of c the choices stay away even further from the optimal arm within the 100 periods of our experiment.

Fig. 8 shows the average offers for 1000 learning direction theory players. We observe the following. First, the initial offers are again around 5. Second, after a steep initial learning effect, the average offers are constant in each treatment.[15] Third, during this initial learning phase, marked differences between the treatments emerge, with the ult10 treatment converging to the optimal minimal offer.

Hence, we see that the situation faced by the players in an ultimatum game is such that players who learn according to these models of boundedly rational behavior would in general not learn

[15] This is not surprising given that this is a discrete Markov process. Notice also that the stationary distributions of offers are independent from the initial guesses.

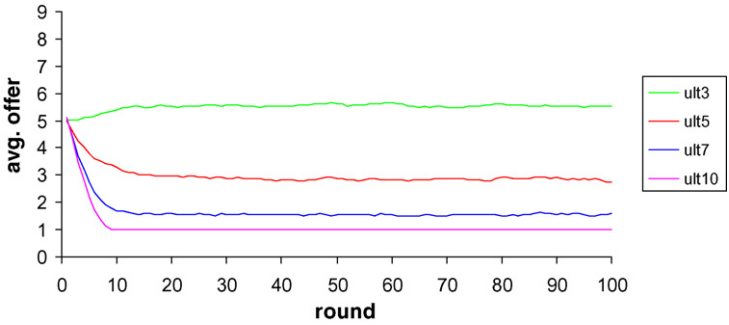

Fig. 8. Theoretical prediction of the average behavior using learning direction theory.

to make the optimal minimal offer either. Moreover, these models of boundedly rational learning display some qualitative similarities with the actual experimental data. This concerns in particular the differences between the treatments and the lack of a clear trend towards the optimal minimal offer.

6. Concluding remarks

The main contribution our paper makes to the literature is providing a characterization of the difficult task faced by the proposing players A in ultimatum games if they were to learn their part of the subgame perfect equilibrium. We conclude that even when the behavior of the receiving players B is qualitatively similar to that of their subgame perfect equilibrium strategy, such that players A face a similar incentive structure, one should not expect players A to converge to the optimal minimal offer unless the players B adhere exactly to the subgame perfect equilibrium strategy of accepting all offers, providing the players A with a noise-free learning environment. This conclusion is based on two elements.

First, we designed a laboratory experiment of a stylized ultimatum game in which we abstract from the coevolutionary aspects of adaptive behavior in a standard ultimatum game experiment, and in which there is also no scope for altruism or fairness considerations. That is, we organized an experiment that basically matches the situation studied in multi-armed bandit problems, framing it as an ultimatum game. The behavior of the receiving players B is fixed from the outset in such a way that making the minimal offer of 1 is optimal. We show that *even* if the learning task for the proposing players A is made as easy as possible while maintaining realistic assumptions concerning the behavior of players B,[16] players A do not really learn this, notwithstanding a learning opportunity of 100 periods (unless the players B behave exactly as in the subgame perfect equilibrium without ever rejecting any offer). Average offers decrease, but very slowly. Hence, the lack of convergence to the subgame perfect equilibrium offer in ultimatum game experiments is not necessarily related to coevolutionary aspects of learning or to fairness considerations.

Second, we put the experimental data into perspective by deriving some theoretical benchmarks from a number of learning models. Analyzing a range of learning theories (from approximately optimal learning based on the Gittins index to boundedly rational learning) in a setup that matches

[16] Leloup (2000) shows that one should not expect convergence to the optimal offer in a situation in which there is almost no difference in expected payoffs of players A for offers between 10 percent and 50 percent of the possible offer range.

exactly the experimental design, we show that the lack of convergence to the optimal minimal offer is an inherent characteristic of the difficult learning task faced by players A, even when the behavior of players B is fixed at a strategy that is close to the subgame perfect equilibrium strategy.

Finally, the theoretical benchmarks suggest some insights into the actual learning behavior of the experimental subjects. Their initial choices suggest that they are reasonably good at taking the structure of the choice situation they face into account (in particular the interdependence of the offers that is included in the adjusted Gittins model, which only includes those hypotheses that are in line with this interdependence, but not in the Gittins model), as their initial choices are consistent with the unbiased guesses of the approximately optimal learning model. But these initial guesses happen to be far away from the real probabilities that actually determine the optimal minimal offer. The weak downward trend in the data suggests that the experimental subjects (for whatever reason) increasingly neglect the initial information about this interdependence between the offers and continue to experiment, as predicted by those learning models that assume independent arms. Although this implies more experimentation than predicted by the model of approximately optimal learning, helping to overcome misleading initial guesses, convergence towards the optimal minimal offer only very partly takes place.

Acknowledgements

We thank Antonio Cabrales, Ido Erev, Steffen Huck, Jon Leland, Paolo Patelli, Robin Pope, Giulio Spelanzon, Massimo Warglien, and two referees for helpful comments, suggestions or discussions. The usual disclaimer applies.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jebo.2004.07.014.

References

Bergemann, D., Välimäki, J., 2001. Stationary multi-choice bandit problems. Journal of Economic Dynamics and Control 25, 1585–1594.

Bolton, G.E., Katok, E., Zwick, R., 1998. Dictator game giving: rules of fairness versus acts of kindness. International Journal of Game Theory 27, 269–299.

Brezzi, M., Lai, T.L., 2002. Optimal learning and experimentation in bandit problems. Journal of Economic Dynamics & Control 27, 87–108.

Cooper, D.J., Feltovich, N., Roth, A.E., Zwick, R., 2003. Relative versus absolute speed of adjustment in strategic environments: responder behavior in ultimatum games. Experimental Economics 6, 181–207.

Forsythe, R.J., Horowitz, J.L., Savin, N.E., Sefton, M., 1994. Fairness in simple bargaining experiments. Games and Economic Behavior 6, 347–369.

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

Gittins, J.C., 1979. Bandit processes and dynamic allocation indices. Journal of the Royal Statistical Society, Series B 41, 148–177.

Gittins, J.C., 1989. Multi-armed Bandit Allocation Indices. John Wiley & Sons.

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

Güth, W., Tietz, R., 1990. Ultimatum bargaining behavior: a survey and comparison of experimental results. Journal of Economic Psychology 11, 417–449.

Kirman, A.P., Vriend, N.J., 1995. Evolving market structure: a model of price dispersion and loyalty (mimeo) Paper presented at the Workshop Economic ALife. Santa Fe Institute.

Kirman, A.P., Vriend, N.J., 2001. Evolving market structure: an ACE model of price dispersion and loyalty. Journal of Economic Dynamics and Control 25, 459–502.

Leloup, B., 2000. May learning explain the ultimatum game paradox? (GRID Working Paper No. 00-03) Ecole Normale Supérieure de Cachan.

Rabin, M., 2000. Risk aversion and expected-utility theory: a calibration theorem. Econometrica 68, 1281–1292.

Roth, A.E., Erev, I., 1995. Learning in extensive-form games: experimental data and simple dynamic models in the intermediate term. Games and Economic Behavior 8, 164–212.

Sarin, R., Vahid, F., 1999. Payoff assessments without probabilities: a simple dynamic model of choice. Games and Economic Behavior 28, 294–309.

Selten, R., Stoecker, R., 1986. End behavior in sequences of finite prisoner's dilemma supergames. A learning theory approach. Journal of Economic Behavior and Organization 7, 47–70.

Thaler, R.H., 1988. The ultimatum game. Journal of Economic Perspectives 2, 195–206.

Vriend, N.J., 1997. Will reasoning improve learning? Economics Letters 55, 9–18.

Appendix: Instructions to the players

Introduction

- This is a decision experiment. The instructions are simple, and if you pay attention, you can gain a reasonable amount of money. From now on till the end of the experiment you are not allowed to communicate with each other. If you have a question, please raise your hand. You are not allowed to use paper, pen, calculator, or any other material not provided by the organizers of the experiment.
- Each of you will play repeatedly the same basic game. Before explaining how often, and with whom you will play this game, we will first explain the basic game as such.

The Basic Game

- There are two players: player A, and player B. Player A has a pie cut in 9 equal slices. Player A makes a proposal to player B concerning the distribution of the pie. Player A can offer to player B from 1 up to 9 slices. Only whole slices are allowed. Player B can do 2 things. First, player B can accept the proposal of player A, which will then be carried out, player B getting the number of slices proposed by player A, and player A keeping the rest of the pie. Second, player B can reject the proposal of player A, in which case the pie perishes immediately, and both players will get nothing.
- Example: if player A proposes to give player B 1 slice, and player B accepts, then player B's payoff will be 1 slice, and player A will keep 8 slices. If, however, player B rejects the proposal, then the payoff for both players will be 0 slices.

The Experiment

- You will play the same basic game for 100 rounds. In each round you will play the role of the proposer: player A. Each round you will be matched `at random' with some player B. You will never play more than once against the same player B.
- The players B with whom you will be matched are drawn from a large population of Artificially Intelligent agents, making their decisions using some computer algorithm.
- Each of those players B's behavior is systematic in the following sense: If a specific player B has accepted an offer x then that player B would have accepted as well any offer greater than x. And if offer x had been rejected by that player B, so would have been all offers smaller than x. Of course, different players B might have different opinions about which offers are acceptable or not. The players B do not change their behavior over time.
- During the experiment, your computer screen will be divided into 2 windows. The upper window will give you general messages, ask for input, etc. The lower window will display the history of your experiment. This window will be scrollable (using the arrows ↑ and ↓), such that you have always access to the complete history. The history will list all rounds, the identity of the specific player B you were matched with, the offer you made, player B's response, and the resulting payoff for you.
- To make your offer, please enter a number. Remember that only integer values from 1 to 9 can be chosen. Please, before pressing Enter, always make sure that you did not make a typing-error.
- There is no time limit for your decisions.

Payment

- You will be paid according to the total payoffs you realized. For each slice of a pie gained you will get . . . . . .[1] At the end of the experiment, we will add up your payoffs, and calculate your monetary rewards. This will be done in a separate room, so you will not see what other players earned.

[1] The monetary reward per slice was specified according to the treatment actually being run.