Experimental Economics, 6:141–158 (2003)

Reinforcement and Directional Learning in the Ultimatum Game with Responder Competition

BRIT GROSSKOPF
Harvard Business School, 174 Baker Library, Soldier's Field Road, Boston, MA 02163, USA
email: bgrosskopf@hbs.edu

Abstract

Demands in the Ultimatum Game in its traditional form with one proposer and one responder are compared with demands in an Ultimatum Game with responder competition. In this modified form one proposer faces three responders who can accept or reject the split of the pie. Initial demands in both ultimatum games are quite similar, however in the course of the experiment, demands in the ultimatum game with responder competition are significantly higher than in the traditional case with repeated random matching. Individual round-to-round changes of choices that are consistent with directional learning are the driving forces behind the differences between the two learning curves and cannot be tracked by an adjustment process in response to accumulated reinforcements. The importance of combining reinforcement and directional learning is addressed. Moreover, learning transfer between the two ultimatum games is analyzed.

Keywords: ultimatum game, competition, reinforcement learning, directional learning, learning transfer

JEL Classification: C72, C91, D83

1. Introduction

In search for robust phenomena guiding individual economic behavior, Reinhard Selten initiated the discussion about a qualitative learning theory in repetitive decision tasks, called directional learning (e.g. Selten and Stoecker, 1986; Selten and Buchta, 1998). The basic idea can be illustrated with an often quoted simple example: Consider a marksman who repeatedly tries to hit the trunk of a tree with a bow and an arrow. After a miss he will have the tendency to aim more to the left if the arrow passed the trunk to the right, and more to the right in the opposite case. This ex-post consideration of choices that might have been better in round t and the according adjustment in round $t + 1$ without taking into account future consequences, induces sequential dependencies in the data. Even though directional learning has been found to successfully describe quite a large body of experimentally observed data, it has often been critizised for its lack of quantitative predictions.[1]

It therefore comes at no surprise that the examination of the growing experimental literature on learning reveals a tendency toward the development of parsimonious models that do not explicitely model sequential dependencies and therefore directional learning (see e.g. Crawford, 1995; Mookherjee and Sopher, 1997; Erev and Rapoport, 1998; Erev and

Roth, 1998a; Slonim and Roth, 1998; Camerer and Ho, 1999). This paper analyzes the extent to which sequential dependencies in the data need to be accounted for.[2]

An interesting example that illustrates the potential discrepancy between the two lines of research is provided by the study of learning in the ultimatum game. In this two-player game, one player (the proposer) is asked to propose a division of a pie, and the second player (the responder) can accept or reject the proposal.[3] Our analysis of the observed round-to-round behavior (see also Mitzkewitz and Nagel, 1993) shows that proposers tend to increase their share (demand) following an acceptance, but reduce it after a rejection, a phenomenon consistent with directional learning. Nevertheless, Roth and Erev (1995) demonstrate that the main results obtained in previous studies of ultimatum and similar games can be accounted for by simple reinforcement learning models that ignore round-to-round directional responses.

The main goal of the current paper is to illustrate that the results obtained by Roth and Erev simply imply that in certain games individual round-to-round dynamics cancel out leaving no effect on the shape of the learning curve. Yet, there are games in which the individual sequential dependencies are likely to be persistent and influence the learning curve. Section 2 presents one example: It is shown that a minimal modification of the ultimatum game changes the impact of sequential dependencies. Without influencing the theoretical predicted equilibria and the reinforcement-based simulated learning curves, the introduction of responder competition leads proposers to increase their demands which is not offset by rejection driven decreases. The importance of combining reinforcement and directional learning is addressed. Moreover, it is shown that the learning transfer between the two ultimatum games heavily depends on which ultimatum game is played first. Section 3 concludes.

2. The experiment

The ultimatum game with responder competition involves one proposer and three responders. If none of the responders accepts (spontaneous response), all receive nothing. A random draw determines which of the accepting responders gets the payoff if more than one responder accept. This is different to the way in which Güth et al. (1997) (GMR) determine the decisive responder in their ultimatum game with 5 responders. The responder that gets the payoff (when there are multiple acceptances) is the one with the lowest acceptance threshold (strategy method). Besides introducing fiercer competition, the experimental design of GMR eliminates all equilibria except the subgame perfect Nash equilibrium.

2.1. Experimental design

The experiment was conducted in 10 sessions (two per day) at the University Pompeu Fabra in Barcelona during spring 1998. Subjects were undergraduates in their first to fourth year studying economics, business or public administration. Analyzing whether the order in which the games are played matters, a so-called Version MT and a Version TM were run. In Version MT, the modified ultimatum game (M) (one proposer and three responders) was played first for six rounds followed by six rounds of the traditional ultimatum game (T) (one proposer and one responder). In Version TM, the traditional ultimatum game (T) was played first for six rounds followed by six rounds of the modified UG (M).

Sixteen undergraduates participated in each session. After entering the lab, participants found two sealed envelopes. They were only allowed to open the first one at the time the instructor first welcomed them. After having finished the first six periods (first part of the experiment) subjects were told to open the second envelope, in which they found the instructions for the next six periods (second part of the experiment).[4]

The hand-run experiment was conducted as a demand game where proposers could state a demand between 0 and 100 (pie size was equal to 100 points = 160 Ptas., at that time valued at approx. 150 Ptas. to $1) divisible by 5. This restriction was imposed in order to facilitate subjects' decision process and to have a higher chance of collecting multiple observations per strategy.[5] People were re-grouped using a random-matching procedure.[6]

During the entire experiment the only information feedback was what a participant had earned in the previous round, i.e. a proposer was only told whether his demand was accepted or not. He was not given any information about the number of accepting/rejecting responders. Respectively, responders were only informed about their own payoff. At the end of the experiment subjects were paid privately according to their accumulated payoffs plus a 500 Ptas. show up fee. Subjects participating in one session were not allowed to participate in another one. Sessions were about 90 minutes in duration; "always proposers" earned on average 1.500 Ptas., "switching proposers" 1.100 Ptas. and "always responders" 700 Ptas (including the show-up fee).

Since 16 subjects were interacting in each experimental session, 8 of them were proposers and 8 responders in the traditional UG. In the modified UG only 4 subjects were proposers but 12 were responders. Those 4 proposers were proposers in both ultimatum games. The other 4 proposers (to make 8 in the traditional UG) changed from either being a proposer in the traditional UG to being a responder in the modified UG (Version TM) or from being a responder in the modified UG to being a proposer in the traditional UG (Version MT). The following analysis does not differentiate between participants who changed their roles and those who did not, because no significant difference in behavior could be found. All results presented in this paper are robust to the inclusion or exclusion of any set of participants.

2.2. Game theoretic predictions

Traditional UG. There are subgame-perfect equilibria (SPE) in which the proposer demands 95 and the responder accepts, and would accept a demand of 100 with probability less than or equal to 0.95. There are SPE in which the proposer demands 100 and the responder accepts with probability greater than or equal to 0.95, and there are SPE in which the proposer mixes between 95 and 100 and the responder accepts demands of 100 with probability exactly 0.95. By contrast, any split of the pie is a Nash equilibrium outcome.

Modified UG. The game theoretic predictions do not change in the modified UG. The SPE are nearly the same, but the cutoff probability for acceptance of 100 demands is lower. So, for both games, there are two pure-strategy SPE and many mixed-strategy SPE. Similarly, the concept of Nash equilibrium permits any split of the pie.

2.3. *Reinforcement learning predictions*

The reinforcement learning model used in Roth and Erev (1995) (RE) can be described as follows. In round $t+1$, player i has the propensity $q_i^k(t+1)$ to play the kth pure behavioral strategy. The strategy chosen by a player together with the strategies of the player(s) with whom he is paired determine the payoff of the players. Payoffs received in round t are used to update players' propensities in round $t+1$. Suppose player i has played his kth behavioral strategy in round t and has received a payoff of $w_i^k(t)$. His propensity to play strategy k in the following round $t+1$ is then updated (by possible depreciation ϕ) as follows

$$q_i^k(t+1) = \phi q_i^k(t) + w_i^k(t). \tag{2.1}$$

The propensity of an unchosen strategy l is updated as

$$q_i^l(t+1) = \phi q_i^l(t). \tag{2.2}$$

The probability $p_i^k(t+1)$ that player i chooses his kth pure strategy in round $t+1$, is given by

$$p_i^k(t+1) = \frac{q_i^k(t+1)}{\sum_{j=1}^n q_i^j(t+1)}, \tag{2.3}$$

where n is the number of possible pure behavioral strategies.[7]

Figure 1 shows the simulated predicted trajectories of mean demands using the simple reinforcement updating mechanism proposed by RE. The simulations are run as follows: A simulated player chooses a demand (proposer) or maximal acceptable demand (responder).[8] The discounting parameter ϕ is set to equal 0.9. Initial propensities (IPs) for proposers and

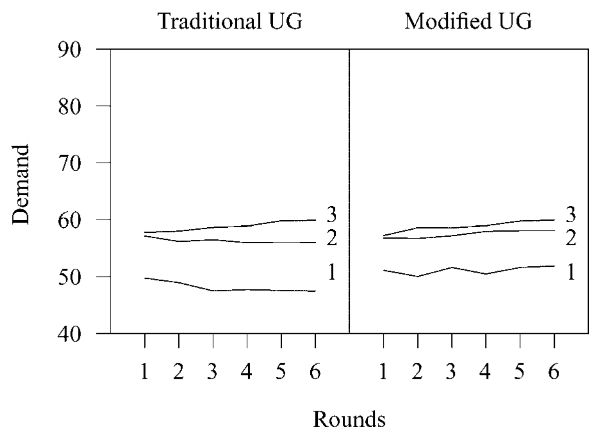


Figure 1. Predicted mean demands. Using the simple reinforcement model for different sets of initial *IP*s, $\phi = 0.9$ (1—uniform *IP*s, 2—RE's *IP*'s, 3—RE's *IP*'s for proposer and responder accept everything).

responders are either assumed to be uniform (each strategy has an equal chance of being chosen) or taken from RE.[9] The strength of the initial propensities, $S(1)$, is set equal to five because it is the mean as well as the median of all pure strategies' payoffs.

Figure 1 clearly suggests that there does not seem to be a significant difference between the predicted behavior in the two versions of the ultimatum game. Even if we simulate responders as to never reject (simulated path 3), we see that proposers' demands stay rather constant and are similar in the the two ultimatum game. However, as we will see in the following section, observed behavior in both ultimatum games is strikingly different.

Why does the simple reinforcement model do so poorly in this game, whereas it could generate behavior resembling observations in the market game (UG with proposer competition) which is very similar to the modified UG?

Public information as a first step towards virtual updating. The public announcement of the selling price after each round in the market game is incorporated in the model of RE as follows: Each player's propensity to play the winning bid is updated, i.e. the reinforcement w(equal to the profit from the winning bid, i.e. 10—winning bid) is distributed among the ten buyers, e.g. if the winning bid is 8, the tendency to play 8 is increased by $\frac{2}{10}$ for all players. Obviously, proposers in the traditional UG are "handicapped" since they do not observe the outcome of any other bargaining pair, whereas buyers in the market game are allowed to search for a "better set" in their gradual groping towards optimal behavior by adapting the winning bid strategy. A "losing" buyer in the market game is thus more likely to increase his bid in the next round, which leads to the rather quick convergence to the subgame-perfect equilibrium. It turns out that this difference is the driving force behind the stickiness of proposers' demands in the traditional UG.

In fact, Al Roth presents an unpublished manuscript (Erev and Roth, unpublished) in the Handbook of Experimental Economics, in which the authors show that reinforcement learning is too slow in minimum action coordination games unless the "winner's" (highest payoff) strategy is reinforced. This is similar to the updating in the market game.

Slonim and Roth (1998) use the reinforcement model with 20% local experimentation that reinforces "similar" strategies to the one played besides a forgetting (or discounting) parameter of $\rho = 0.001$ $(q = 1 - \rho)$ to explain observed behavior in ultimatum games with varying stake sizes. They suggest that the lower rejection frequency is the reason that the proposers in the higher stakes' conditions of the ultimatum game learn to make higher demands. From Figure 1 we see that even in the case when responders accept everything, demands do not increase, in fact with local experimentation simulated play of the two ultimatum game needs at least 100 rounds to diverge.

No payoff–no learning. This condition holds in the updating of the propensities of proposers in the traditional UG but is violated in the updating done in the market game due to the informational advantage discussed before.[10]

Thus, the divergence in simulated behavior in the market game and the traditional UG is driven by the different information available to the players.

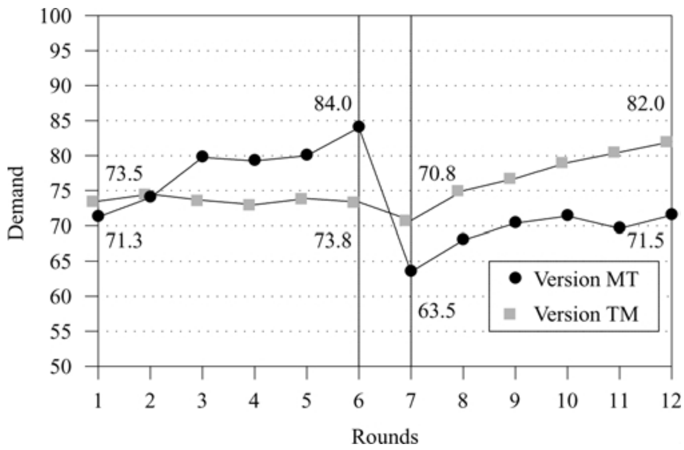


Figure 2. Mean demands in version MT. (Modified UG → Traditional UG) and TM (Traditional UG → Modified UG.)

2.4. *Experimental results*

Figure 2 shows the average demand over time in the two different ultimatum games in Version MT and Version TM, respectively. Studying this figure reveals most of the interesting features of this experiment (see figures 6 and 7 in the Appendix for all individual demands and rejections).

Observation 1 (Ex Ante Equivalence). Initial behavior in the two ultimatum games does not differ significantly.

The observed distributions of demands in the first round of each version are not significantly different from each other (Kolmogorof–Smirnov test, $p = 0.985$). The mean demand in the modified UG of Version MT is 71.3 in the first round, which is only slightly below the mean demand in the traditional UG of Version TM which is 73.5. Intuitively, this suggests that ex ante both ultimatum games are perceived to be similar by the proposers. Everything that happens after the first round can therefore be attributed to the experience players gather while interacting in the experiment.

Interestingly, Observation 1 can also be interpreted as an underestimation of the importance of the decision rule (see Messick et al., 1997). Proposers do not seem to differentiate between the traditional UG, where they need the acceptance of the only one responder and the modified UG, where they only need one acceptance out of three.

Observation 2 (Competition). Demands in the modified UG increase steadily, and are higher than in the traditional UG.

Figure 2 shows that the mean demand in the modified UG increases from 71.3 to 84 in Version MT and from 70.8 to 82 in Version TM. The mere existence of three responders drives demands closer to the subgame perfect equilibrium prediction. The non-parametric

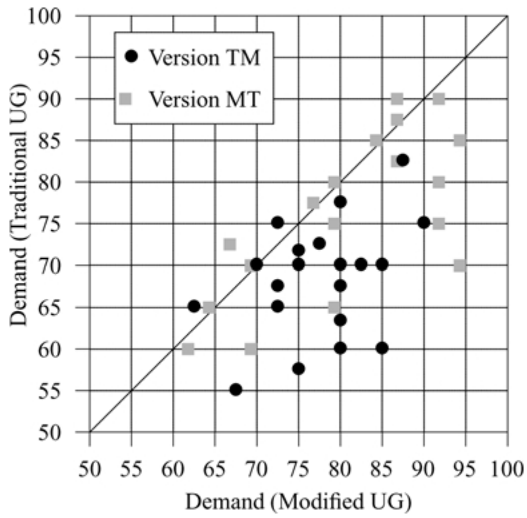


Figure 3. Comparison of median demands in the traditional and in the modified UG. (Each data point plots the median demand (pooled over all six rounds) of a proposer in the modified UG against the median demand (pooled over all six rounds) of the same proposer in the traditional UG.)

Page test confirms that median demands increase in the modified UG over time ($p < 0.001$ for both versions). Mean demands in the traditional UG when it is played first stay rather constant (73.5 to 73.8). The slope of an estimated regression line is not significantly different from zero (see Appendix, Table 2). In Version TM, mean demands in the traditional UG increase from 63.5 to 71.5. This uncharacteristic increase is driven by an initial "over-reaction" of proposers changing from the modified UG to the traditional UG, which we will discuss in more detail later (see Observation 4).

Figure 3 plots the median demand of each proposer in the modified UG against the median demand of the same proposer in the traditional UG. Although there are a few observations above the 45°-line, most proposers demand more in the modified UG than in the traditional UG.[11] Clearly, introducing responder competition leads proposers to behave differently. To see why this is so, consider what a proposer fears the most—a rejection. In the modified UG (given that responders' behavior does not change[12]) the probability that a certain demand is rejected decreases. A proposer receives a payoff of zero only when *all* responders reject the demand. Observation 3 quantifies the difference in responders' behavior.

Observation 3 (Rejection rates). Demands in the traditional UG are rejected more often than demands in the modified UG.

Table 1shows the relative rejection frequencies in the two ultimatum games. An important distinction is made between "effective" and "actual" rejections. We observe an "effective" rejection, if all responders with whom the proposer is paired, reject. Obviously, the great difference between the two UGs lies in the effective relative rejection frequencies. Demands are rejected 7 times (35 times!) more often in the traditional UG than in the modified UG

Table 1. Total relative rejection frequencies.[13]

Modified UG		Traditional UG
Actual	Effective	Actual = Effective
Version MT		
21.67%	4.17%	30.00%
Version TM		
12.78%	0.83%	29.17%

of Version MT (TM). Is this effect driven solely by the statistical decrease in the rejection probability in the modified UG or do responders behave differently?

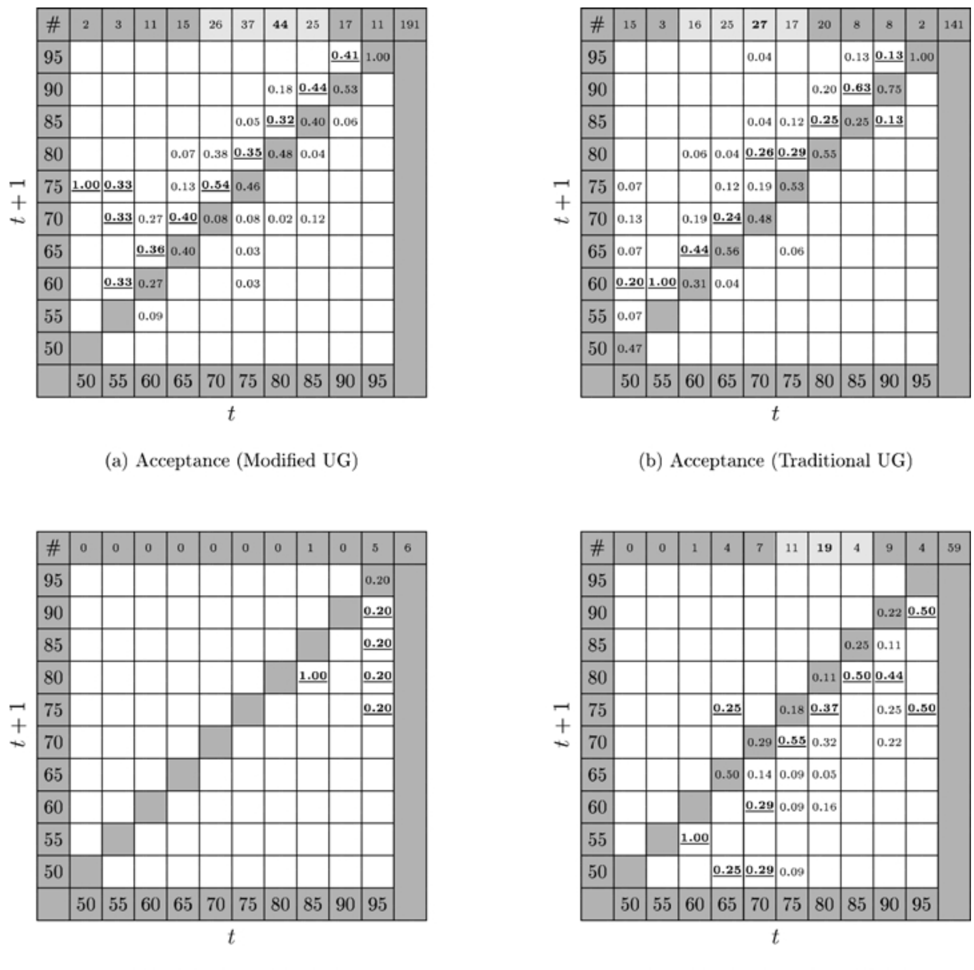
In both versions, a one-tailed test of equality of proportions, using the normal approximation to the binomial distribution (see Glasnapp and Poggio, 1985), confirms that the actual relative rejection frequency is significantly higher in the traditional UG than in the modified UG ($p < 0.05$).[14] To control for proportionally equivalent demands in the traditional and the modified UG we run a robust logit regression (details are specified in the Appendix, Table 3). We find that the higher the demand, the more likely is a rejection in both types of ultimatum games. There is no difference per se between the modified and the traditional UG. In later rounds a demand is more likely to be accepted in both ultimatum games, however, less so in the modified UG.

Responders seem to reject when they know that their punishment will be effective. Bolton and Zwick (1995) have shown that when punishment by rejection is less effective, demands are higher and rejections are less frequent. The strategy method used by GMR is more appropriate for analyzing responders' behavior. They find that acceptance thresholds are not really different in the very first round but drastically decrease over time. However, this strong effect is also driven by the fact that the decisive responder, i.e. the accepting responder who receives the payoff, in GMR was the one with the lowest acceptance threshold, so that responders were in an even more competitive environment.

In fact, equity-based fairness models (e.g. Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000) predict higher demands in the modified UG compared to the traditional UG and lower rejection rates in the former. As with other static predictions, we see that these models fail to predict first period behavior which is not significantly different across games, but could potentially be seen as an approximation of behavior of experienced subjects.

The individual round-to-round dynamics are best described in a transition matrix. Figure 4 shows the proportion of proposers who, after having experienced an acceptance (rejection) of the proposed split of the pie, decide to increase, decrease or to leave their proposal unchanged. The underlined bold numbers are off-diagonal modal relative frequencies, the vertical axis refers to the demand that was accepted (rejected) and the horizontal axis identifies the decision in the following period. The top row of each matrix shows the total number of demands.

Take for example the matrix entry (70, 80) in the transition matrix of acceptance in the modified UG (figure 4(a)). It indicates that 38% of the proposers, whose demand of 70 was accepted, increased their demand in the next round to 80. We also see that a total of 26 demands of 70 were accepted. Consider further the entry (95, 75) in the rejection matrix of



(a) Acceptance (Modified UG)

(b) Acceptance (Traditional UG)

(c) Rejection (Modified UG)

(d) Rejection (Traditional UG)

Figure 4. Transition matrix of demands aggregated over version MT and version TM in case of acceptance and rejection. (Columns indicate demands in period t, rows correspond to demands in $t + 1$, i.e. an entry in the acceptance matrix denotes the percentage of proposers (whose demand was accepted) that decreased or increased their demands or left it unchanged. Respectively in the rejection matrix: underlined bold numbers are modal off-diagonal relative frequencies, the top row shows the total number of demands of the corresponding demand level in round t. Lighter shaded boxes resemble "quasi-boxplots.")[15]

the traditional UG (figure 4(d)). We find that 50% of the proposers, whose demand of 95 was rejected, decrease their next period demand to 75. We can also see that a total of four demands of 95 were rejected.

These observations are in line with directional learning (DL) (Selten and Stoecker, 1986; Selten and Buchta, 1998). The basic idea behind this qualitative reasoning process is that, given the behavior of others, a player considers ex-post whether he *could* have obtained a higher payoff by choosing a different strategy. If he changes his strategy it should be in the direction of potentially higher payoffs. Or, at least, more choices should go into the "right" than into the wrong direction.

To reject predictions of DL we need to see more modal off-diagonal relative frequencies below the diagonal in the acceptance case and above in the rejection case. Looking at the four matrices we clearly see that this is not the case. Most proposers either do not change their demand or increase it after the previous one was accepted. If the previous demand was rejected, the majority of proposers demand less in the next round. "Successful" proposers tend to go up to the next highest level of demand. Some even jump to a higher level than the adjacent one.

The ex-post bounded-rationality assumption of directional learning seems to explain individual adaptive behavior.[16] We can also see that proposers in the modified UG are not stopped by rejections in their upwards movement, since rejections in the modified UG are almost non-existent. Moreover, demands in the modified UG concentrate at a higher level than demands in the traditional UG. Note that both transition matrices in the acceptance case show the same individual behavioral pattern indicating similar individual behavioral dynamics (sequential dependencies) in both ultimatum games.

Conducting separate χ^2-tests for each demand level gives statistical assurance.[17] For all demand levels but 70, we cannot reject the null hypothesis of having equal transition behavior in the traditional and the modified UG. Proposers who chose 70 in the modified UG are significantly more likely to increase their demand in the next round than are proposers in the traditional UG that demanded 70 ($p < 0.01$). A comparison in the rejection case is not possible due to a lack of observations in the modified UG. But clearly, aggregated behavior of proposers is characterized by very similar round to round responses of proposers, but shaped by the difference in effective relative rejection frequencies.[18]

Clearly, the payoff for following the directional learning rule is very different in the two ultimatum games. To change one's choice in line with DL predictions is not particularly reinforcing in the traditional UG, but is very reinforcing in the modified UG. The obtained average payoff from moves in line with DL is 79.35 (79.56) in the modified UG when it is played first (second) as opposed to 54.80 (43.07) in the traditional UG when it is played first (second). Comparing these obtained average payoffs to the average payoff that proposers would have gotten if they *had not changed* their choice following an acceptance highlights the ex-post superiority of changing one's choice.

Proposers in the modified UG would have earned only 69.27 (70.20) on average if they had not changed, whereas proposers in the traditional UG would have done better if they had not changed their choices (69.27 and 73.06 respectively).

Observation 4 (Learning Transfer). Proposers changing from the modified UG to the traditional UG (Version MT) decrease their demands immediately and drastically. However, proposers changing from the traditional UG to the modified UG (Version TM) do not change their demands significantly.

Modified UG $\longrightarrow$ Traditional UG (version MT). Looking at the change in behavior from round 6 to round 7, we are interested in whether demands that once came close to the subgame-perfect equilibrium can persist. This turns out not to be the case. Changing from the modified UG to the traditional UG induces proposers to *immediately* and *drastically* decrease their demands. Comparing the mean demand in the last round of the modified UG with the first round of the traditional UG we find a decrease by 20.5 points (i.e. a drop from 84 to 63.5).[19] Note that this is not due to any rejections experienced by proposers in the 6th round of the modified UG since every single demand (in each of the conducted sessions) in the last round of the modified UG was accepted. One could ask whether this is driven by differences in the behavior of proposers who were proposers in the modified UG and those who changed from being a responder in the modified UG to being a proposer in the traditional UG. Our hypothesis was that these "new" proposers could be more concerned with fairness issues after having experienced the less favorable side of the market. Even though we find slightly higher demands of "always" proposers (65.5) in comparison to "new" proposers (61.5), the difference turned out not to be significant (Kolmogorof–Smirnov test, $p = 0.12$).

Thus, it seems as if proposers understood the schematic principle (the decision rules). By round 6 of the modified UG proposers may have realized why they were able to increase their demands significantly and that the "new" game (starting in round 7) is to some extent similar, however with a decreased bargaining power. However, proposers seem to "over-adjust," at least initially. Demands in the first round of the traditional UG when it is played after the modified UG are not only lower than last round demands in the modified UG but are also significantly lower than demands in the traditional UG when it is played first (compare 73.5 with 63.5, Kolmogorof–Smirnov test, $p = 0.014$). Figure 5 plots the average payoff by proposed demands. By pooling all demands in the first six rounds, the aim of the exercise was to find out what would have been the payoff-maximizing choice (given the empirical distribution). The open cirles indicate the expected payoff-maximizing choice, 70 (with 80 following suit) in the traditional UG and 90 in the modified UG. Clearly, by decreasing demands as far down as 63.5, proposers "over-adjust" when changing from the modified UG to the traditional UG, maybe even negatively transfer what they had experienced so far.

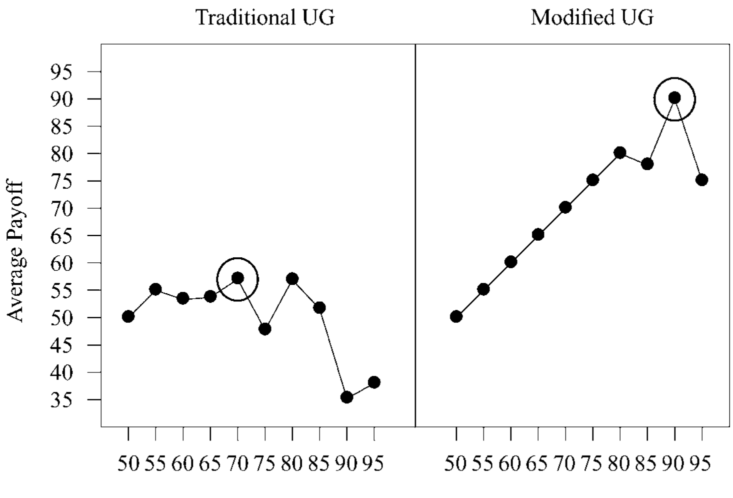


Figure 5. Average payoffs by proposed demands. (Pooled data of the first 6 rounds).

Traditional UG → Modified UG (version TM). On the contrary, proposers experiencing a change from the traditional UG to the modified UG do not change their behavior in the first period of the new treatment significantly (compare 73.8 with 70.8). It is only a period later that they slightly start to demand more. It seems as if these proposers "under-adjust" when changing from the traditional UG to the modified UG. This seems to manifest the underestimation of the decision rule, as observed by Messick et al. (1997). Why do proposers take their time to recognize their strengthened bargaining power? Is loss-aversion driving this unchanged behavior? Or is it rather the incapability to assess the complexity of the new environment? Facing three (instead of only one) responders demands more reasoning on the proposers' side. They might simplify their decision problem by keeping their behavior unchanged and waiting for further feedback. It is not clear which explanation is the driving force.[20]

The question about transfers arises. Psychologists find that transfer is most likely to occur when situations are perceived to be quite similar and that subjects often have to be reminded of the similarity between situations.[21] Note that the experimental procedure treated both ultimatum games separately, so it was left to the subjects to deduce any common (or different) features.

Taking all observations together, we see that the observed learning curve in the traditional UG is rather flat and, as shown in figure 1, can be simulated rather successfully with a reinforcement based adaptation mechanism that ignores sequential dependencies. However, we observe a significant increase in demands over time in the modified UG that cannot be tracked by an adjustment process in response to accumulated reinforcements. Sequential dependencies in the data, offset in the traditional UG, play an important role in the modified UG and shape the slope of the learning curve. In the modified UG the "upward" trend of demands is not offset by rejection driven decreases, since only six times out of 191 a demand of a proposer was rejected. On the contrary, rejections in the traditional UG are almost ten times as high. Clearly, the "upward" and "downward" trend interact leading to a rather "stable" aggregate demand pattern in the traditional UG. Therefore, ignoring sequential dependencies can mislead predictions. Grosskopf (2001) suggests a way to account for these dependencies by incorporating an updating procedure of unchosen strategies in the framework of the simple reinforcement model. Camerer et al. (2002) also update unchosen strategies in their modified experience-weighted-attraction model. Another approach, suggested by Erev and Roth (1998b), might be to model directional learning as a cognitive strategy that players are able to select besides stage game strategies. Cognitive strategies can be task-specific, but still, the individual adaptation process can be described as "a relative permanent change in behavioral potentiality that occurs as a result of reinforced practice" (Kimble, 1961). Barron and Erev (2001) propose a learning model that allows for learning among cognitive strategies instead of stage-game strategies alone. Further research in this area is needed to evaluate the different approaches and result in better quantification.

3. Discussion

This paper reports on an experiment that modifies the simple rules of a game without affecting the equilibrium predictions. It is shown that a minimal modification of the ultimatum game changes the observed learning curve. The mere existence of three responders (who can accept or reject) leads proposers to increase their demands significantly.

By varying the order in which the two ultimatum games are played, we are able to analyze learning transfer. When changing from the modified UG to the traditional UG, proposers seem to "over-adjust". They drastically and immediately decrease their demands, while proposers changing from the traditional UG to the modified UG seem to "under-adjust" by not changing their behavior at all.

Examination of the individual behavior of proposers reveals that this striking difference between the modified and the traditional UG prevails even though individual sequential dependencies in the data are observed to be the same in the two ultimatum games. Proposers tend to demand more after a previously stated demand was accepted and less if it was rejected. This pattern can be observed in both types of ultimatum games. However, increases in demands after an acceptance seem to be canceled out by decreases after rejections in the traditional UG. Since rejections are far less frequent in the modified UG, the increase in demands prevails. Pure reinforcement (without adding further parameters) could not have predicted this strikingly different pattern, since it ignores the fact that changing one's choice after receiving a positive payoff (e.g. after an acceptance) might yield a much higher payoff than just sticking to the previous choice. We have shown that these sequential dependencies are shaping the aggregate behavior. Clearly, even though there are games in which directional learning dynamics offset each other (e.g. traditional UG), there are certainly other games (not necessarily very different) in which sequential dependencies are likely to influence behavior (e.g. modified UG). Ignoring this possibility might lead to misspecified models. Therefore, a combination of the the two approaches seems potentially worthwhile. This paper does not attempt to propose a quantitative solution. It is meant as an experimental approach to address the seeming disagreement between reinforcement and directional learning.

Directional learning implicitly grants players the "feeling of not being entirely satisfied with their choices", even when they receive a positive payoff. It seems reasonable to assume that players are likely to evaluate ex-post the potential performance of other strategies. This idea of improving upon one's choice might be more descriptive of human behavior than the basic stimuli-response framework of pure reinforcement. Extending the pure reinforcement model through directional reasoning might allow for a better modeling of bounded rational but intelligent agents in different classes of similar games and serve as a crucial step towards a deeper understanding of cognition driven human behavior.

Appendix: Robust regression results

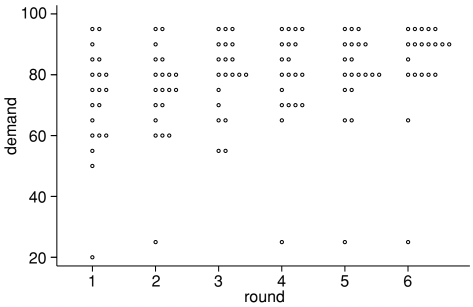
We run robust regressions with the data of all ultimatum games considered pairwise. Observations of the ultimatum game mentioned first in each column were assigned the dummy $UG = 1$ for each pair. Observations in each experimental sessions were blocked in order to account for the interdependence of those observations.

$$demand = a + b_{UG}UG + b_{round}round + b_{inter}UG \cdot round,$$

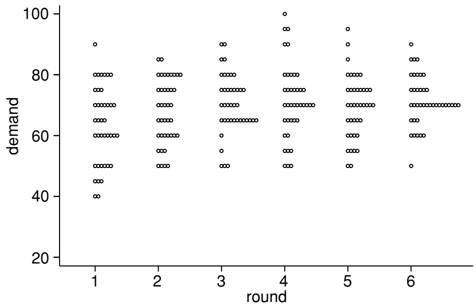
Table 2. Robust regression results.

	Mod_{MT} Trad_{TM}	Mod_{TM} Trad_{MT}	Trad_{MT} Trad_{TM}	Mod_{MT} Mod_{TM}
Intercept (a)	73.96**	64.48**	69.83**	64.48**
	(41.31)	(38.04)	(46.08)	(38.04)
b_{round}	−0.09	1.31**	2.14**	1.31**
	(−0.20)	(3.24)	(5.76)	(3.24)
b_{UG}	−4.04	5.36**	4.13*	5.44
	(−1.06)	(2.36)	(1.76)	(1.44)
b_{inter}	2.41**	0.83	−2.23**	1.01
	(2.49)	(1.51)	(−3.85)	(1.06)
adj. R^2	0.05	0.19	0.06	0.14

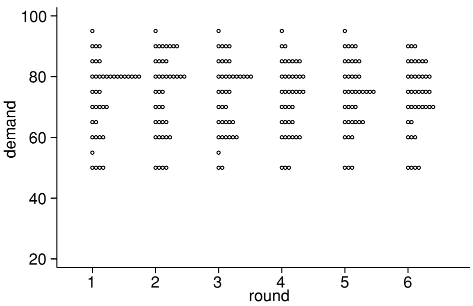
$^*p < 0.1$, $^{**}p < 0.05$, t-values are given in parenthesis.



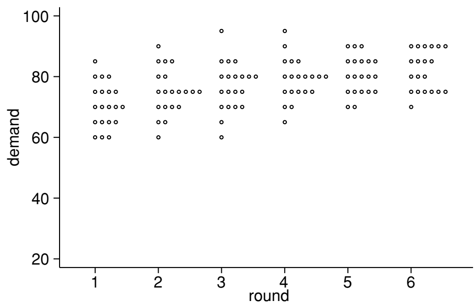
(A) MODIFIED UG (VERSION MT)



(B) TRADITIONAL UG (VERSION MT)



(C) TRADITIONAL UG (VERSION TM)



(D) MODIFIED UG (VERSION TM)

Figure 6. Dotplots of the different ultimatum games.

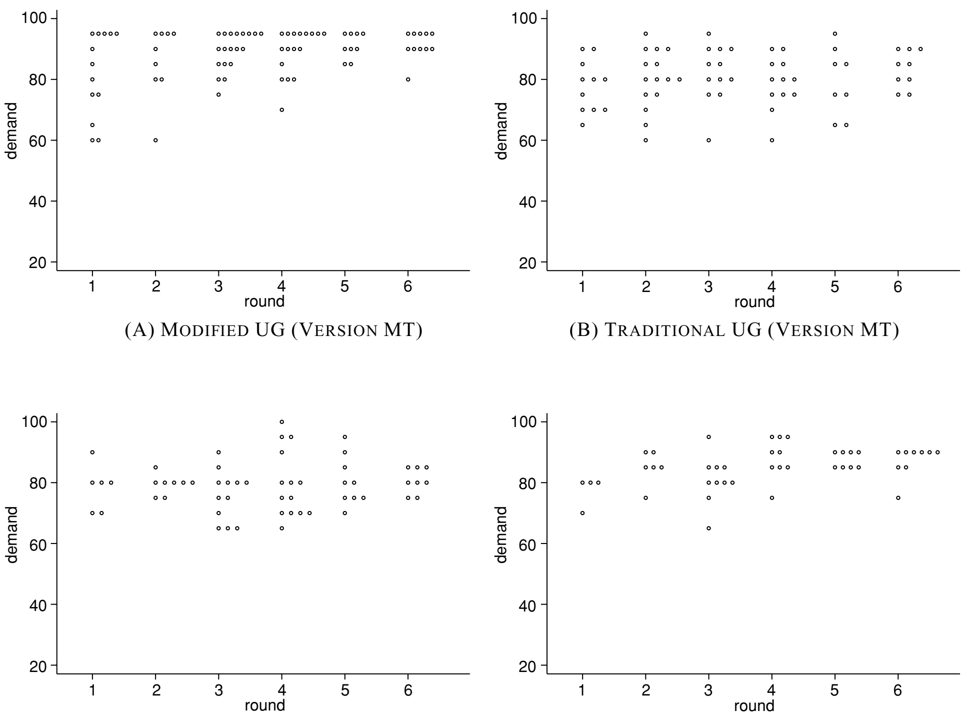
Table 3. Robust logit regression results.

Parameter		p-value
Intercept (a)	11.65	0.000
b_{dem}	−0.15	0.000
b_{UG}	0.72	0.068
b_{round}	0.47	0.000
b_{inter}	−0.15	0.000
# observations	581	
Log Likelihood	−242.68	

where *round* is a round index and *inter* is the interaction between rounds and the type of UG in order to account for potential differences in the learning rate.

Robust logit regression results

We run a robust logit regression with the data of the traditional and modified UG when they are played first, where $response = 1$ if the demand was accepted and 0



(A) MODIFIED UG (VERSION MT)

(B) TRADITIONAL UG (VERSION MT)

(C) TRADITIONAL UG (VERSION TM)

(D) MODIFIED UG (VERSION TM)

Figure 7. Dotplots of the individual rejected demands (actual rejections, not necessarily effective ones).

otherwise.

$$response = f(a + b_{dem}demand + b_{UG}UG + b_{round}round + b_{inter}UG \cdot round),$$

where $f(x) = \frac{1}{1+e^x}$, and *demand* is the demand of the proposer, $UG = 1$ if the data is from the modified UG and 0 otherwise, *round* is a round index and *inter* is the interaction between rounds and the type of UG in order to account for potential differences in the learning rate.

Acknowledgments

Financial support from the Spanish Ministry of Education and Science under contract PB94-0663-C03-02 is gratefully acknowledged. I thank Yoella Bereby-Meyer, Gary Bolton, Antonio Cabrales, Gary Charness, Ido Erev, Nick Feltovich, Atanasios Mitropoulos, Albert Satorra and especially Rosemarie Nagel for helpful comments and suggestions.

Notes

1. Various studies successfully apply directional learning, see e.g. Mitzkewitz and Nagel (1993) for the ultimatum game, Kuon (1994) for bargaining, Cachon and Camerer (1996) for coordination games, Cason and Friedman (1997) for markets, Nagel and Tang (1998) for the centipede game, Kagel and Levin (1999) for auctions, Nagel and Vriend (1999) for oligopoly markets and Grosskopf et al. (2001) for individual decision making tasks. Studies that find mixed success for directional learning are for example, Huck et al. (1999) for oligopoly markets, Ho et al. (1998) for coordination games and Duffy and Nagel (1997) for max-guessing games.
2. Recently, a new wave of models started to realize this point. We will come back to them at the end of this paper.
3. First studied by Güth et al. (1982), it has been the subject of many experimental studies. For a summary of the bargaining literature see the corresponding chapter by Alvin E. Roth in the *Handbook of Experimental Economics* or the survey by Bolton (1998).
4. Each set of instructions was self-contained and did not refer to the other set.
5. These simplified games thus have finite strategy spaces. This feature was chosen to offer computational manageability without over-simplifying the game.
6. Subjects were told that they would face different “people” (proposers, responders) during those six rounds. Proposers in the traditional UG never encountered the same responder twice whereas proposers in the modified UG were never confronted with the same set of three responders.
7. Various probability functions have been used in previous research, including a logit or exponential form, power form and probit form. Each have advantages and disadvantages. The choice of probability function in this paper follows RE.
8. The game was simplified so that choices were integers between 1 and 9. This was done in close resemblance of RE. For representation purposes the numbers were multiplied by 10. Figure 1 shows aggregated results of 200 simulated quasi-experiments.
9. Aggregated over all four countries. Calculations of IPs are taken from RE.
10. Suppose a proposer demanded 8 in the ultimatum game which was rejected, leaving him with a payoff of zero. His strategy to play 8 gets reinforced with a payoff of zero (which is equivalent to no reinforcement at all). Has the proposer “learned” anything from such a “negative” experience? According to the simple reinforcement model he has not, i.e. a proposer starts the next round as if he had never played the one before, which is probably not very realistic.
11. Conducting a Wilcoxon signed ranks test on the session-level data shows that demands in the modified UG’s are significantly higher than demands in the traditional UG’s ($p < 0.01$).
12. It is not straightforward that responders’ acceptance threshold decreases since the expected loss of a rejection is less in the three responders case. Following Gale et al. (1995) this would actually mean that responders’ behavior is noisy, i.e. responders would be expected to reject more often. However, results from the impunity

game of Bolton and Zwick (1995) would suggest that we observe less rejections since the effectiveness of a rejection decreases.

13. Taken from the entire sequence of games regardless of the size of demands which were rejected. See Table 3 (in the Appendix) for logit regression results.
14. The specific test statistic is $Z = (p_1 - p_2) = S_{p_c}$, where p_i is the proportion of demands rejected in subsample i. S_{p_c} is an estimate of the standard error of $p_1 - p_2$, $S_{p_c} = \{p_c(1 - p_c)[(1/N_1) + (1/N_2)]\}^{0.5}$. p_c is an estimate of population proportion under the null hypothesis of equal proportions, $p_c = (p_1N_1 + p_2N_2)/(N_1 + N_2)$, where N_i is the total number of demands in subsample i.
15. One proposer in the modified UG of Version MT decreased his demand in round 4 from 80 to 25 without having experienced a rejection. He kept on demanding 25 in the remaining rounds of the modified UG. That is why there are only 197 instead of 200 possible data points.
16. Mitzkewitz and Nagel (1993) already noted in their incomplete information setting of the traditional UG that an increase in demands is more likely to happen after an acceptance and a decrease in demands will probably occur after a rejection.
17. In order to fulfil the criteria for conducting a χ^2-test, adjacent strategies are combined such that demands in round $t + 1$ could fall into two groups, less or equal than demands in round t or higher than demands in round t.
18. In fact, Abbink et al. (2001) distinguish between a pure learning based explanation for observed behavior in the ultimatum game and an explanation that is entirely driven by the taste for punishment of the responders. They note that any difference in proposer behavior across treatments must have its origin in different responder behavior.
19. The Wilcoxon signed ranks test of individual demands in the 6th round of the modified UG and the 1st round of the traditional UG (7th round of total play) shows that this is a significant difference ($p < 0.01$).
20. Contrary to the modified UG of Version MT there are rejections in the last round of the traditional UG of Version TM which also may have an influence on behavior in round 7.
21. See Gick and Holyoak (1980), (1983), Perkins and Solomon (1988) and Solomon and Perkins (1989).

References

Abbink, K., Bolton, G.E., Sadrieh, A., and Tang, F.-F. (2001). "Adaptive Learning versus Punishment in Ultimatum Bargaining." *Games and Economic Behavior*. 37(1), 1–25.

Barron, G. and Erev, I. (2001). "On Adaptation, Maximization, and Reinforcement Learning Among Cognitive Strategies." Mimeo, Technion, Haifa.

Bolton, G.E. (1998). "Strong and Weak Equity Effects: Evidence, Significance and Origins." In D. Budescu, I. E. and R. Zwick (eds.), *Games and Human Behavior, Essays in Honor of Amnon Rapoport*, Hillsdale, NJ.

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

Bolton, G.E. and Zwick, R. (1995). "Anonymity versus Punishment in Ultimatum Bargaining." *Games and Economic Behavior*. 10, 95–121.

Cachon, G.P. and Camerer, C. (1996). "The Sunk Cost Fallacy, Forward Induction and Behavior in Coordination Games." *Quarterly Journal of Economics*. 111, 165–194.

Camerer, C.F. and Ho, T.H. (1999). "Experience—Weighted Attraction Learning in Games: A Unifying Approach." *Econometrica*. 67, 827–874.

Camerer, C.F., Ho, T.H., and Wang, X. (2002). "Individual Differences in EWA Learning with Partial Payoff Information." Mimeo.

Cason, T. and Friedman, D. (1997). "Price Formation in Single Call Markets." *Econometrica*. 65(2), 311–345.

Crawford, V.P. (1995). "Adaptive Dynamics in Coordination Games." *Econometrica*. 63, 103–143.

Duffy, J. and Nagel, R. (1997). "On the Robustness of Behavior in Experimental Beauty-Contest Games." *Economic Journal*. 107, 1684–1700.

Erev, I. and Rapoport, A. (1998). "Coordination, 'Magic,' and Reinforcement Learning in a Market Entry Game." *Games and Economic Behavior*. 23, 146–175.

Erev, I. and Roth, A. (Unpublished). "Adaptation and Imitation in Experimental Games." Technical Report.

Erev, I. and Roth, A.E. (1998a). "Predicting How People Play Games: Reinforcing Learning in Experimental Games with Unique, Mixed Strategy Equilibria." *American Economic Review*. 88(4), 848–881.

Erev, I. and Roth, A.E. (1998b). "On the Role of Reinforcement Learning." In D. Budescu, I.E. and R. Zwick (eds.), *Games and Human Behavior, Essays in Honor of Amnon Rapoport*, Hillsdale, NJ.

Fehr, E. and Schmidt, K. (1999). "A Theory of Fairness, Competition and Cooperation." *Quarterly Journal of Economics*. 114, 817–868.

Gale, J., Binmore, K.G., and Samuelson, L. (1995). "Learning to be Imperfect: The Ultimatum Game." *Games and Economic Behavior*. 8, 56–90.

Gick, M.L. and Holyoak, K.J. (1980). "Analogical Problem Solving." *Cognitive Psychology*. 12, 306–355.

Gick, M.L. and Holyoak, K.J. (1983). "Schema Induction and Analogical Transfer." *Cognitive Psychology*. 15, 1–38.

Glassnapp, D. and Poggio, J. (1985). *Essentials of Statistical Analysis for the Behavioral Sciences*. Columbus, OH: Merrill.

Grosskopf, B. (2001). "Merging Reinforcement and Directional Learning—A Generalized Model." Mimeo.

Grosskopf, B., Erev, I., and Yechiam, E. (2001). "Foregone with the Wind." Mimeo.

Güth, W., Marchand, N., and Rulliere, J.-L. (1997). "On the Reliability of Reciprocal Fairness—An Experimental Study." Mimeo.

Güth, W., Schmittberger, R., and Schwarze, B. (1982). "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior and Organization*. 3, 367–388.

Ho, T., Camerer, C., and Weigelt, K. (1998). "Iterated Dominance and Iterated Best Response in Experimental p-Beauty Contests." *American Economic Review*. 88(4), 947–969.

Huck, S., Normann, H.-T., and Oechssler, J. (1999). "Learning in Cournot-Oligopoly—An Experiment." *Economic Journal*. 109(454), C80–95.

Kagel, J. and Levin, D. (1999). "Common Value Auctions with Insider Information." *Econometrica*. 67(5), 1219–1238.

Kimble, G.A. (1961). *Hilgard and Marquis' Conditioning and Learning*. Englewood Cliffs, Prentice Hall.

Kuon, B. (1994). *Two-person Bargaining Experiments with Incomplete Information*. Springer Lecture Notes in Economics and Mathematical Systems (No. 412). Berlin, Heidelberg, New York: Springer.

Messick, D.M., Moore, D.A., and Bazerman, M.H. (1997). "Ultimatum Bargaining with a Group: Underestimating the Importance of the Decision Rule." *Organizational Behavior and Human Decision Processes*. 69, 87–101.

Mitzkewitz, M. and Nagel, R. (1993). "Experimental Results on Ultimatum Games with Incomplete Information." *International Journal of Game Theory*. 22, 171–198.

Mookherjee, D. and Sopher, B. (1997). "Learning and Decision Costs in Experimental Constant Sum Games." *Games and Economic Behavior*. 19, 97–132.

Nagel, R. and Tang, F.F. (1998). "Experimental Results on the Centipede Game in Normal Form: An Investigation on Learning." *Journal of Mathematical Psychology*. 42, 356–384.

Nagel, R. and Vriend, N. (1999). "An Experimental Study of Adaptive Behavior in an Oligopolistic Market Game." *Journal of Evolutionary Economics*. 9, 27–65.

Perkins, D.N. and Salomon, G. (1988). "Teaching for Transfer." *Educational Leadership*. 46, 22–32.

Roth, A. and Erev, I. (1995). "Learning in Extensive Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term." *Games and Economics Behavior*. 8, 164–212.

Roth, A.E. (1995). "Bargaining Experiments." In J. Kagel and A.E. Roth (eds.), *Handbook of Experimental Economics*, Princeton University Press.

Salomon, G. and Perkins, D.N. (1989). "Rocky Roads to Transfer: Rethinking Mechanisms of a Neglected Phenomenon." *Education Psychologist*. 24, 113–142.

Selten, R. and Buchta, J. (1998). "Experimental Sealed Bid First Price Auction with Directly Observed Bid Functions." In D. Budescu, I.E. and R. Zwick (eds.), *Games and Human Behavior, Essays in Honor of Amnon Rapoport*, Hillsdale, NJ.

Selten, R. and Stoecker, R. (1986). "End Behavior in Sequences of Finite Prisoner's Dilemma Supergames: A Learning Theory Approach." *Journal of Economic Behavior and Organization*. 7(1), 47–70.

Slonim, R. and Roth, A.E. (1998). "Financial Incentives and Learning in Ultimatum and Market Games: An Experiment in the Slovak Republic." *Econometrica*. 66(3), 569–596.