

## Relative versus Absolute Speed of Adjustment in Strategic Environments: Responder Behavior in Ultimatum Games

DAVID J. COOPER  
*Case Western Reserve University*

NICK FELTOVICH  
*University of Houston*

ALVIN E. ROTH  
*Harvard University*

RAMI ZWICK  
*Hong Kong University of Science and Technology*

### ***Abstract***

Learning models predict that the relative speed at which players in a game adjust their behavior has a critical influence on long term behavior. In an ultimatum game, the prediction is that proposers learn not to make small offers faster than responders learn not to reject them. We experimentally test whether relative speed of learning has the predicted effect, by manipulating the amount of experience accumulated by proposers and responders. The experiment allows the predicted learning by responders to be observed, for the first time.

**Keywords:** Game Theory, learning, bargaining

**JEL Classification:** C7, C9, D83

### **1. Introduction**

There's an old joke about two men awakened by a bear while camping in the woods. One starts putting on his running shoes, and the other says, "A man can't outrun a bear." The first man responds, "I don't have to outrun the *bear*, I just have to outrun *you*."

The point is that *relative* speed sometimes matters as much as absolute speed. This is also a robust prediction of models of learning in strategic environments. How a player learns to adjust his behavior, and how quickly, depends on what the other players are doing, and on how quickly they are changing their behavior.

The advantage of learning theories over other kinds of theories is clearest for games in which players' behavior changes a great deal over time, but does so slowly. For other kinds of games, namely games in which behavior very quickly converges to some stable behavior, or (especially) games in which players' behavior changes little over time, it is easy to see

the appeal of theories that de-emphasize learning. The aim of the present paper is to show that even in games of this latter kind, in which behavior appears to change only slowly if at all, the relative speed of learning may play a critical role in determining what behavior is observed.

Roth and Erev (1995) showed that very simple learning models could qualitatively track experimentally observed behavior both in games in which behavior quickly converged to a (perfect) equilibrium and in games in which it did not, for games with similar equilibrium predictions. Erev and Roth (1998) found that closely related learning models could predict behavior in real time, i.e. they could approximate the speed and magnitude of players' adjustments in their behavior as they gained experience with games. The success of the learning models in these papers rests not so much on their ability to explain behavior in any single game, but rather on their ability to provide a unified theory of behavior across a wide variety of games that induce vastly different behavior.

Perhaps the most surprising result in Roth and Erev (1995) is the explanation of behavior in the ultimatum game through the use of a learning model. This simple bargaining game is among the most studied games in experimental economics.<sup>1</sup> It is a two player game between a "proposer," who proposes how to divide a fixed sum between the two players, and a "responder," who either accepts the proposed division, in which case the proposed split is implemented, or rejects it, in which case both players earn zero. When both players evaluate outcomes only in terms of their own payoff, the perfect equilibrium of this game is for the proposer to offer the smallest feasible positive amount to the responder, and for the responder to accept. The experimental results, in contrast, consistently show that small offers are made rarely, and are frequently rejected when made. The most common outcome is that the proposer offers the responder something in the range of 40–50%, and the responder accepts.

Reinforcement models of learning explain the robust experimental results in terms of the relative speed of learning that the game induces between the two kinds of players.<sup>2</sup> These theories predict that a responder who receives a very small offer adjusts his behavior little, after either accepting or rejecting it, because (since it is a small offer) accepting it gives only slightly greater reinforcement than rejecting it. However a proposer who makes a moderate offer and has it accepted reaps a large reward, unlike a proposer who makes a small offer and has it rejected. Thus proposers learn not to make small offers much faster than responders learn not to reject them. And, once proposers stop making small offers (or make them very rarely), there is little further opportunity for responders to learn not to reject them. So if players begin with somewhat diffuse propensities over what offers to make and accept, the learning dynamics reproduce the experimentally observed behavior.

However, there exists little experimental evidence that the behavior of responders changes over time. This is not entirely unexpected, since the learning models predict that learning by responders will be very much slower than that of proposers, but it presents a critical empirical challenge to learning theories. These theories posit that the responders and proposers learn in *precisely the same way*, even though the game causes them to learn at different speeds. While learning models appear to explain a wide range of strategic behavior in very different games,<sup>3</sup> the failure to detect responder learning in ultimatum games raises the question of whether this success is somehow accidental. Indeed, alternative theories have been presented

for the ultimatum game, in which players have unchanging preferences over distributions of outcomes rather than just their own payoffs (a “taste for fairness,”), that also capture the main features of ultimatum game data.<sup>4</sup> Unlike learning theories, these theories predict that the behavior of responders will not change over time (see e.g. Ochs and Roth, 1988; Bolton, 1991; Bolton and Ockenfels, 2000; Fehr and Schmidt, 1999; Rabin 1993).<sup>5</sup> In some ways this approach seems more natural than learning theories, since responders in the ultimatum game are not faced with any strategic complexities that can only slowly be learned. However since quite a range of learning theories also predict the experimental results, the existing data cannot help determine if the conceptual explanation of responder behavior embodied in the learning models has predictive power beyond that of alternative theories.

The present paper reports an experiment designed to directly test the learning theory predictions about responder behavior in ultimatum games, by examining if the relative speeds of learning by proposers and responders have the predicted effect. The design of the experiment rests heavily on the learning model. That is, the design is intended to make it easy to see learning on the part of responders if it arises in just the way that the learning model predicts, as a function of the simultaneous learning being experienced by the proposers.

Specifically, we will seek to directly influence the relative speeds of learning by varying the amount of experience that proposers and responders obtain. In one condition of our experiment, called the  $1 \times 1$  condition, responders and proposers will play the ultimatum game equally often. In the other condition, called the  $2 \times 1$  condition, responders will play twice as often as proposers. For example, a responder playing his tenth ultimatum game will be receiving a proposal from a proposer playing his fifth game. The reinforcement learning model predicts that, because a responder in the  $2 \times 1$  condition is playing less experienced proposers, he will receive more low offers, and learn more quickly to accept them, compared to the  $1 \times 1$  condition in which responders and proposers acquire equal experience. Therefore rejection rates should be lower in the  $2 \times 1$  treatment.<sup>6</sup>

The present experiment is designed so that, if the learning models are correct about the kind of learning going on in ultimatum games, it will have enhanced power to detect the learning of responders. The experimental results are consistent with the predictions of the learning model. As predicted by the learning model, rejection rates are significantly lower in the  $2 \times 1$  treatment than in the  $1 \times 1$  treatment. Beyond predicting a treatment effect, the learning model also predicts a mechanism underlying the treatment effect. Responders in the  $2 \times 1$  treatment should receive more low offers, and therefore have more opportunities to learn to accept low offers. When we econometrically control for the proportion of low offers received by the responders, we find that the treatment effect vanishes. The results of our experiment thus find the predicted effect, for the predicted reason.

In summary, the behavior of responders in the ultimatum game presents a test of the apparent generality of theories of learning in games. Alternative explanations of observed behavior in ultimatum games have been proposed that do not involve any learning, but rather suppose that responders have unchanging preferences for fairness. Since existing experimental results are consistent with both sorts of theories, we have designed an experiment that should make it easy to detect learning by responders if they are learning in exactly the same fashion as proposers. Our experimental results are consistent with the predictions of reinforcement learning models. In particular, we can predict changes in responders'

behavior as we manipulate the relative speeds of learning by proposers and responders. Thus the learning models, which explain the easily observed features of the ultimatum game in terms of the relative speed of learning on each side of the game, also give us the ability to predict even very subtle effects having to do with the relative speed of learning.

## 2. Experimental methods

We ran sixteen experimental sessions, consisting of eight  $1 \times 1$  sessions (with equal numbers of proposers and responders) and eight  $2 \times 1$  sessions (with twice as many proposers as responders). A total of 250 subjects participated in the sixteen sessions, split into 112 subjects in  $1 \times 1$  sessions and 138 subjects in  $2 \times 1$  sessions. Subjects were primarily undergraduates at the University of Pittsburgh. No subject participated in more than one session.

In each session, subjects were randomly assigned to computer terminals; the assigned terminal determined the role of the subject (proposer or responder). Subjects remained in the same role throughout a session. Subjects were handed written instructions and given time to read them. The instructions were then read aloud by a monitor in order for the rules of the game to be common knowledge. Subjects' questions were also answered at this time.

Most sessions consisted of 50 periods of play. One session was cut short after 40 periods due to a computer crash. Two sessions that finished early were run for an additional 10 periods in order to observe the further development of play. Subjects were not told in advance how many periods would be played, but they did know that the session would last no longer than 2 hours.

The sequence of events in a period was as follows. First, proposers and responders were randomly matched. The matching algorithm prohibited players from being matched to the same counterpart in two successive periods; otherwise, all matchings were equally likely. Matchings were anonymous. Next, proposers were prompted to choose an offer. Offers were constrained to be whole dollar amounts between \$1.00 and \$10.00 inclusive (out of \$10.00). After offers were made and verified, each responder was shown the offer of his proposer and prompted to choose a response (accept or reject). After responses were made and verified, players were shown the result of play in that period (own action, other player's action, own payoff and other's payoff) and prompted to press a key to continue. After all subjects had pressed the key, the next period began. Each subject's computer screen kept track of his or her recent history of play.

At the end of a session, one period was randomly chosen and subjects earned their payoff (in dollars) for that period in addition to a show-up fee of \$5.00.<sup>7</sup>

In the  $1 \times 1$  condition, all responders and proposers played in every period. In the  $2 \times 1$  condition, all responders played every period while proposers only played every other period. More specifically, proposers were split into two groups with one group playing only odd numbered periods and the other playing only even numbered periods. This treatment was designed to make the responders relatively more experienced than the proposers. For instance, when a responder plays his tenth ultimatum game he is matched with a proposer playing only his fifth game.

To maximize the likelihood that any observed differences in subject behavior in the two cells were due only to changes in the relative frequency of play, we made it difficult for proposers in the  $2 \times 1$  sessions to realize that they were playing in only every second period. The instructions did not make any mention of period numbers, so we could unobtrusively vary the number of periods that different subjects played, without the use of deception. All references to “period” numbers on a subject’s screen were in the subject’s own time (so that in a  $2 \times 1$  session, what was period 20 to a responder would be period 10 to a proposer).

### 3. The reinforcement learning model: Description and predictions

In this section, we use simulations based on a reinforcement learning model to generate predictions for this experiment. More precisely, each individual will be modeled as a reinforcement learner, and the predictions of the learning model for the current experiment will be developed by running computational simulations of the experiment in which simulated players will be matched as in the actual experiment. Our goal is not to find the learning model that best fits the data from this experiment, but rather to develop a simple model that allows us to robustly predict outcomes for this (and other) experiments.

To maintain comparability with the analysis in Roth and Erev (1995), we use one of their simple models for the simulations.<sup>8</sup> The technical details of the reinforcement learning model and the simulations are contained in Appendix A of this paper. To understand the gist of the material, however, a brief intuitive description of the model and a summary of how it was implemented for the simulations will suffice. The reinforcement learning model formalizes two basic psychological principles. First, strategies that do better are played more frequently over time. Second, the rate of adjustment slows down over time. These two principles, known respectively as the “Law of Effect” (Thorndike, 1898) and the “Power Law of Practice” (Blackburn, 1936), have been validated by numerous experiments.

To formalize these two “laws,” the reinforcement learning model assumes that players place a weight, known as a propensity, upon each of the available strategies. The probability of a strategy being chosen is proportional to its propensity. The central feature of the model is how the propensities are updated following each play. In the simplest version of the model, the propensity for the strategy that was just used is updated (reinforced) by adding the realized payoff to the propensity while the propensities for other strategies are unaffected. We study a version of the model that includes “forgetting”—all propensities are discounted by a fixed factor prior to the updating. Intuitively, this modification to the model puts greater weight on relatively recent experiences.<sup>9</sup>

Because the amount of reinforcement a strategy receives after being played depends on the realized payoff, strategies that yield higher expected payoffs will also receive greater reinforcements. It follows that the reinforcement learning model obeys the Law of Effect. As long as the rate of forgetting is not too great, the total sum of propensities will grow over time. Since the size of payoffs (and hence reinforcements) is not changing over time, it follows that the reinforcement learning model will follow the Power Law of Practice with learning curves becoming flatter over time.

In implementing the reinforcement learning model for the simulations, we mimic the experiments as much as possible. Each (simulated) proposer is allowed to use the ten

offers available in the actual experiment. For simplicity, we restrict responders to the use of cutoff strategies, where a responder's cutoff specifies the lowest offer that play is willing to accept.<sup>10</sup> Higher cutoffs correspond to tougher behavior by responders. The restriction to cutoff strategies gives the responders ten available strategies, like the proposers. For all simulations there were 10 responders. For simulations of the  $1 \times 1$  treatment there were ten proposers, with this doubled to twenty proposers for simulations of the  $2 \times 1$  treatment. These numbers are similar in scale to the numbers of subjects in our sessions. As in the actual experiments, the simulated proposers play in every period for the  $1 \times 1$  simulations and alternate periods in the  $2 \times 1$  simulations. Simulated subjects were randomly matched as in the experiments. All of the simulations were run for 50 periods, the modal number of periods in the actual experiments. This allows us only 25 periods to compare proposers' behavior between the two treatments.

Because our goal is to predict behavior in our experiments, we did not engage in a statistical exercise to find the best possible set of parameters either for these experiments or some previously published experiment.<sup>11</sup> Instead, we restrict the initial values of the propensities to generate a fixed distribution of strategies similar to those typically observed in ultimatum game experiments. We also pick plausible values for the strength of the initial propensities and the rate of forgetting, the parameters that govern the speed of learning, to serve as a baseline.<sup>12</sup> After generating predictions for the baseline parameter values, we then vary the strength of the initial propensities and the rate of forgetting to study the robustness of these predictions. For all sets of parameters we ran 10,000 simulations, so our predictions are unlikely to depend on the particulars of what random numbers were drawn for the simulations.

The top two panels of figure 1 summarize the results of the baseline simulations while the bottom two panels compare the predicted treatment effects across a variety of parameter values.

The top left panel in figure 1 shows the average offer over the 10,000 baseline simulations for each treatment. The periods shown on the x-axis are given from the proposer's point of view. Thus, the 50th period of a  $2 \times 1$  simulation is shown as the 25th proposer-period, since it is only the 25th time the proposers have made a decision. For both treatments, the average offer rises over time. This increase does not reflect some general theoretical property of the reinforcement learning model, but instead depends on the initial propensities. Since the initial offers are slightly lower on average than the best response to responders' initial behavior, the average offer must rise over time in the simulations. Comparing the plots for the two treatments, the average offers are less in the  $2 \times 1$  treatment than in the  $1 \times 1$  treatment. This difference develops gradually over time.

The top right panel of figure 1 shows the average cutoff over the 10,000 baseline simulations for each treatment. For both treatments, the average cutoff falls over time. This is a general property of the reinforcement learning model. Play in the reinforcement learning model on average moves towards better responses, so it follows that cutoffs must on average be falling. This implies that, holding the offer fixed, rejection rates must decline over time. The magnitude of this decrease is determined by the initial propensities. Comparing the plots for the two treatments, the average cutoff is less in the  $2 \times 1$  treatment than in the  $1 \times 1$  treatment. Lower cutoffs imply lower rejection rates in the  $2 \times 1$  treatment. The difference emerges slowly over time.

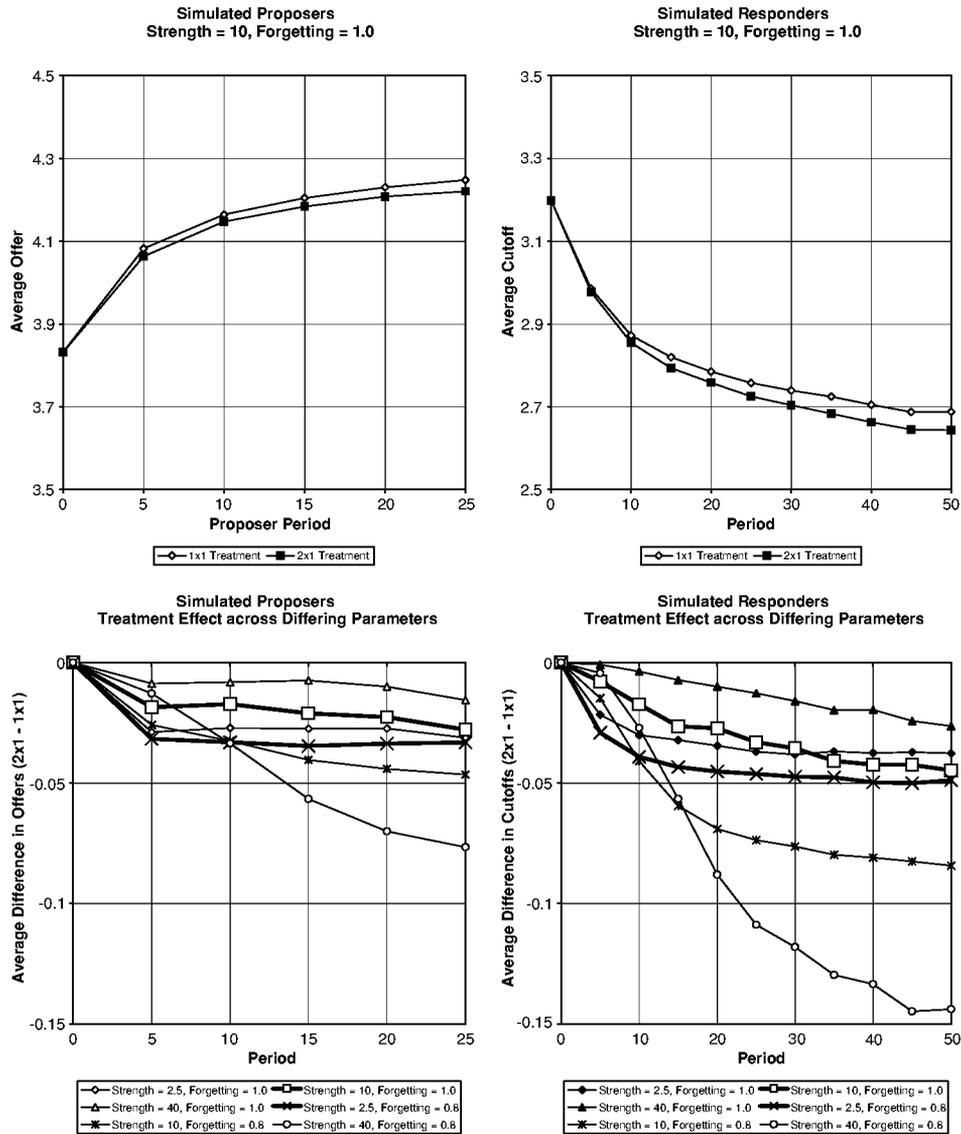


Figure 1. Simulation results—Comparison of treatments.

Based on the baseline simulations, we predict lower offers and lower rejection rates in the  $2 \times 1$  treatment. To explore the robustness of these predictions and the speed with which differences might emerge, we vary the strength of initial propensities and the forgetting parameter. We examine five parameter combinations in addition to the baseline values.<sup>13</sup> The bottom left panel of figure 1 shows the difference in average offers between the two

treatments in these six sets of simulations. Specifically, the graph shows the average offer in the  $2 \times 1$  treatment simulations minus the average offer in the  $1 \times 1$  treatment simulations. As previously, the period number gives the number of times the proposers have made a decision reflecting the different timing for proposers in  $2 \times 1$  sessions. The plot for the baseline simulations is highlighted with large squares and bold lines. Across parameter values the average offer is consistently lower in the  $2 \times 1$  treatment. The size of this treatment effect and the speed with which it emerges vary considerably across parameter values. Thus, we can make a robust prediction that average offers will be lower in the  $2 \times 1$  treatment, but cannot make any strong predictions about how fast this difference will emerge or how large it will be.

The bottom right panel of figure 1 compares the difference in average cutoffs between the two treatments across the six sets of simulations, displaying the average cutoff in the  $2 \times 1$  treatment simulations minus the average cutoff in the  $1 \times 1$  treatment simulations. The plot for the baseline simulations is again highlighted with large squares and bold lines. Across parameter values the average cutoff is consistently lower in the  $2 \times 1$  treatment. The size of this treatment effect and the speed with which it emerges vary considerably across parameter values. For example, consider the plot highlighted with large x's and bold lines. The simulations that generate this plot have both a lower strength for the initial propensities and a lower rate of forgetting than the baseline simulations.<sup>14</sup> After 50 periods, the predicted treatment effect for responders with these parameters is almost identical to the predicted treatment effect for the baseline simulations. However, the effect is much faster to emerge; the plot is almost completely flat after the first ten periods. Econometrically, we would expect to be able to pick up the difference between the two treatments but not necessarily the widening of this effect. More generally, we can make a robust prediction that average cutoffs will be lower in the  $2 \times 1$  treatment, but cannot predict how fast this difference will emerge or how large it will be.

To summarize, the learning model robustly predicts that both average offers and average cutoffs are lower in the  $2 \times 1$  treatment. We therefore predict that (controlling for the number of proposer periods) lower offers should be observed in the  $2 \times 1$  treatment, and (holding the offer fixed) rejection rates should be lower in the  $2 \times 1$  treatment. The model does not make any specific predictions about the magnitude of the treatment effect for either role, but in the simulations it is typically small. Given the subtlety of the predicted treatment effects, little reason exists to expect effects that are obvious to the naked eye. The learning model robustly predicts that the treatment effect for both roles will increase over time, but this increase may be so slight as to be virtually undetectable.

The intuition underlying the treatment effects predicted by the reinforcement learning model is roughly the same for either role. Responders observe offers rising over time. Even ignoring any treatment effects on the proposers, the  $2 \times 1$  treatment makes this increase seem half as fast to the responders. Since they are receiving lower offers, responders in the  $2 \times 1$  treatment learn to use lower cutoffs than responders in the  $1 \times 1$  treatment. From the point of view of proposers, rejection rates decline over time. Even without any treatment effects for responders, the  $2 \times 1$  treatment makes this decline appear twice as rapid to the proposers. Since they are receiving fewer rejections, proposers in the  $2 \times 1$  treatment learn to make lower offers than proposers in the  $1 \times 1$  treatment (or rather, do not learn

as quickly to avoid making low offers). While these predictions have been developed for a specific formulation of Roth and Erev's learning model, the same intuition will hold for any reinforcement learning model that obeys the Law of Effect and the Power Law of Practice.

#### 4. Experimental results

##### 4.1. An overview of the data

Table 1 summarizes the experimental data, broken down by session type. There are relatively few offers of 6 or more, so these offers have been pooled together in a single category. Likewise, given the small number of offers of 1 or 2, these two offers have been pooled into a single category. We report the data both in terms of raw counts and as frequencies. Information on rejections is reported in parentheses.

Before examining Table 1 in detail, a cautionary note is in order. The statistics reported in this table are aggregated over multiple sessions. This aggregation introduces biases into the raw statistics.

For example, the observed decrease in rejection rates does not allow us to automatically conclude that responders are learning to accept more offers. In later periods of the

Table 1. Distribution of proposals and responses by session type.

Offer	Periods 1–15		Periods 16–60		Total	
	Raw data	Frequency	Raw data	Frequency	Raw data	Frequency
<i>1 × 1 Sessions</i>						
1–2	62	.074	69	.033	131	.045
(Rejections)	(51)	(.823)	(50)	(.725)	(101)	(.771)
3	138	.164	286	.137	424	.145
(Rejections)	(75)	(.543)	(93)	(.325)	(168)	(.396)
4	377	.449	1004	.480	1381	.471
(Rejections)	(85)	(.225)	(179)	(.178)	(264)	(.191)
5	232	.276	675	.323	907	.310
(Rejections)	(1)	(.004)	(45)	(.067)	(46)	(.051)
6–10	31	.037	56	.027	87	.030
(Rejections)	(0)	(.000)	(0)	(.000)	(0)	(.000)
<i>2 × 1 Sessions</i>						
1–2	45	.065	62	.040	107	.040
(Rejections)	(39)	(.867)	(49)	(.790)	(88)	(.822)
3	117	.170	261	.169	378	.169
(Rejections)	(67)	(.573)	(126)	(.483)	(193)	(.511)
4	283	.410	716	.465	999	.465
(Rejections)	(41)	(.145)	(109)	(.152)	(150)	(.150)
5	217	.314	451	.293	668	.293
(Rejections)	(2)	(.009)	(0)	(.000)	(2)	(.003)
6–10	28	.041	50	.032	78	.035
(Rejections)	(2)	(.071)	(3)	(.060)	(5)	(.064)

experiment, the distribution of offers is endogenous. If learning by proposers moves them towards offers with higher expected payoffs, there will be negative correlation between the initial probability that low offers are rejected in a session and the probability that such offers are observed in later periods. It follows that low offers in later periods are more likely to come from sessions in which low offers were more often accepted. The resulting aggregation effect biases the observed change in rejection rates for low offers downwards. To control for individual effects (and the resulting aggregation effects), we examine the behavior of responders using probit analysis with a random effects specification.<sup>15</sup>

Pooling data from all treatments over all periods, the average offer is 4.13. The average offer rises slightly over time, changing from an average of 4.04 for the first fifteen periods to an average offer of 4.16 in the remaining periods. The distribution of offers tightens somewhat over time. This is reflected by a fall in the standard deviation of offers, from 1.09 in the first fifteen periods to .97 in the remaining periods.

Average offers are lower in  $2 \times 1$  sessions (4.10 over all periods) than in  $1 \times 1$  sessions (4.14 over all periods). This difference is due primarily to later periods; for periods 16–60, the average offer is 4.20 for  $1 \times 1$  sessions and 4.11 for  $2 \times 1$  sessions. The modal offer is 4 for both types of session in all periods.

Turning to responders, we observe frequent rejection of positive offers. The overall rejection rate is 19.7%, and the modal offer, 4, has a rejection rate of 17.4%. Considering the offers for which there are significant amounts of rejection, rejection rates fall over time. This effect is especially strong for low offers (3 and less). As noted above, we shouldn't read too much into these declines since they may be due solely to aggregation bias.

The response data do not reveal an obvious systematic difference between  $1 \times 1$  and  $2 \times 1$  sessions. Pooling all periods, the rejection rate is slightly lower for offers of 4 or more in  $2 \times 1$  sessions, but substantially higher for offers of 3 or less. These differences must be interpreted with great caution. Not only are there strong individual effects in the data, but the aggregation effects are also stronger for  $1 \times 1$  sessions than for  $2 \times 1$  sessions.

Overall, the data are consistent with observations from earlier ultimatum game experiments.<sup>16</sup> The experiments do not yield the subgame perfect outcome, and are not obviously trending toward the subgame perfect outcome over time. Positive offers are persistently rejected with substantial probability.

#### 4.2. *Econometric analysis of responder data*

Our econometric analysis of the data begins by examining the responder data. We show that there exists a strong treatment effect, providing evidence in favor of learning by responders, and explore the factors that drive this treatment effect. In this section we provide a summary of the analysis. Technical details of how the variables were constructed and how the regressions were run are contained in Appendix B of this paper.

All of our econometric analysis of responders' data uses probit regressions with a random effects specification (to control for individual effects). The dependent variable is the responder's choice, with positive parameters corresponding to a higher probability of rejection. Statistical tests of significance for individual parameter estimates are always two-tailed z-tests, and tests of joint significance are always loglikelihood ratio tests.

Several of our regressions include measures of past behavior by proposers. We use three variables to measure proposers' past behavior: the proportion of previous offers greater than or equal to 5 (high offers), the proportion of previous offers less than or equal to 3 (low offers), and the lagged offer from the preceding period. Since we are interested in learning effects, our analysis includes controls for changes in time. We use a non-linear specification for time, with the variable "Late Periods" being a dummy for observations after the 15th period.

The probit analysis of responders' behavior is summarized in Table 2. Model 1 is a baseline regression that only includes a constant, the current offer, and the dummy for late periods. As we would expect, the current offer achieves a high level of statistical (and economic) significance. If we couldn't detect the negative correlation between current offers and rejection rates, there would be little reason to believe any of the other results. We cannot detect any learning effects here as the dummy for late periods fails to be statistically significant at any standard level. This result is consistent with the results of earlier experiments that have failed to detect learning by responders. As noted in the simulation section, learning over time will be difficult to detect econometrically for subjects whose learning curves flatten out quickly (obeying the Power Law of Practice). The primary purpose of our experiment is to give us an indirect method for detecting learning by responders.

Model 2 adds a dummy for the  $2 \times 1$  sessions, as well as an interaction term between the  $2 \times 1$  dummy and the late periods dummy. The  $2 \times 1$  dummy is easily statistically

Table 2. Probit regressions on responder data (102 subjects, 5058 observations).

Variable	Model 1	Model 2	Model 3	Model 4
Constant	4.892** (.105)	5.009** (.125)	5.090** (.294)	4.974** (.307)
Current offer	-1.468** (.015)	-1.472** (.015)	-1.485** (.016)	-1.467** (.016)
$2 \times 1$ Dummy		-.506** (.098)		.112 (.096)
Late periods	-.048 (.044)	-.019 (.059)	-.075 (.046)	-.104 <sup>+</sup> (.059)
Late periods $\times$ $2 \times 1$ dummy		-.026 (.093)		.054 (.096)
Proportion of high offers (offer $\geq$ 5)			.201 (.136)	.277* (.133)
Proportion of low offers (offer $\leq$ 3)			-1.608** (.147)	-1.504** (.167)
Lagged offer			.020 (.039)	.023 (.040)
Log likelihood	-1315.099	-1307.280	-1291.721	-1291.140

<sup>+</sup>Significantly different from 0 at the 10% level.

\*Significantly different from 0 at the 5% level.

\*\*Significantly different from 0 at the 1% level.

significant at the 1% level, but the interaction term fails to be statistically significant at any standard level. The treatment effect emerges rapidly, with little change in later periods. This again is consistent with our observation from the simulations that a treatment effect should exist regardless of the parameters for the model, but changes over time may be hard to detect for subjects whose learning curves flatten out quickly. Holding all else equal, the impact of moving from the  $1 \times 1$  treatment to the  $2 \times 1$  treatment is a decrease in rejection rates of 9%.<sup>17</sup> It may seem surprising that this non-negligible effect cannot be detected with the naked eye. The magnitude of the individual effects in the responders' data cannot be overstated. Without some control rejection rates, for the individual effects, these easily overwhelm any treatment effect.

Finding the predicted effect for the  $2 \times 1$  treatment does not mean that the theory underlying the prediction is necessarily correct. The reinforcement learning model predicts an effect through the treatment's impact on the distribution of offers that responders observe. Models 3 and 4 allow us to establish that the differences in responders' behavior between the two treatments are indeed driven by differences in the offers that are being observed by responders.

Model 3 introduces our three measures of past behavior by proposers, the proportion of previous high offers (offer  $\geq 5$ ), the proportion of previous low offers (offer  $\leq 3$ ), and the lagged offer from the preceding period. The proportion of low offers easily achieves statistical significance at the 1% level, while the other two measures of past proposer behavior have no statistically significant impact on rejection rates. In terms of economic significance, a 10% change in the proportion of low offers has about a 3% effect on rejection rates holding all else equal. The negative relationship between the proportion of low offers and rejection rates is extremely robust and can even be seen with the naked eye. Take the first 25 periods of data (usually the first half of the experiment) and group subjects into thirds by the proportion of low offers they have observed. We can then calculate each subject's rejection rate for the modal offer of 4 in the remainder of the experiment. Going from the third with the most low offers to the third with the least low offers, the respective rejection rates are 15%, 17%, and 25%. Beyond the current offer, the proportion of low offers in previous periods is by far the most important factor driving rejection rates. No other factor is even close.

With the addition of the three measures of past proposer behavior we can detect a weak decrease in rejection rates over time. The dummy for late periods just barely misses significance at the 10% level, and if we eliminate the non-significant parameters for the proportion of high offers and the lagged offer from the regression, then the dummy for late periods becomes statistically significant at the 10% level. Thus, while we find strong evidence that responders adjust their behavior in reaction to the offers they receive, we still only find weak direct evidence of learning (holding the distribution of offers fixed). This serves to reinforce our earlier point that learning, even when it is quite strong initially, can be difficult to detect directly if it quickly dies out.

Model 4 combines the two variables designed to capture treatment effects with the three measures of past behavior by proposers. Neither the  $2 \times 1$  dummy nor the interaction between the  $2 \times 1$  dummy and the dummy for late periods is statistically significant at any standard level. These two variables also fail to achieve joint significance ( $\chi^2 = 0.81$ ,  $p > .10$ , 2 d.f.). The proportion of low offers remains significant at the 1% level, and the ratio

of high offers edges up to statistical significance at the 5% level. The dummy for late periods achieves statistical significance at the 10% level. Once we control for the past behavior by proposers, the treatment effect observed in Model 2 vanishes. This implies that the treatment effect in Model 2 is driven by responders' reactions to the differing distributions of offers for the two treatments, consistent with the explanation of the treatment effect provided by a reinforcement learning model.

Overall, our analysis of the responders' data yields strong conclusions. As predicted by the reinforcement learning model, we are able to detect lower rejection rates in the  $2 \times 1$  treatment. These differences are closely tied to differences between the distribution of offers observed in the two treatments. In particular, responders react strongly to differences in the proportion of low offers received in previous periods.

#### 4.3. *Econometric analysis of proposer data*

This section summarizes our analysis of proposer data. Once again, see Appendix 2 of this paper for more technical details.

Our econometric analysis of proposer data uses ordered probit regressions with a random effects specification (to control for individual effects). All regressions on proposer data measure time from the perspective of the proposer. Thus, the tenth period of a  $2 \times 1$  session is considered equivalent to the fifth period of a  $1 \times 1$  session, since in both cases the proposer is playing for the fifth time. We refer to the "proposer period" to make it clear that we are referring to the number of times a proposer has played. As with responders, we use a non-linear specification for time with proposers. We subdivide proposer periods into three classes: proposer periods 1–15, proposer periods 16–25, and proposer periods 26–60. We refer to these as early, middle, and late proposer periods respectively. The break following 25 proposer periods isolates proposer periods that only contain data from  $1 \times 1$  sessions. This guarantees that any differences we identify between the treatments aren't due solely to the fact that there are twice as many proposer periods in the  $1 \times 1$  sessions.

Table 3 summarizes the ordered probit regressions on proposer data. Model 1 is a baseline regression that only includes the dummies for middle proposer periods and late proposer periods. Both of these terms are positive and significant at the 1% level, indicating a statistically significant increase in offers over time. The predicted shifts are small in magnitude. The average offer is predicted to rise by about twelve cents between the early and middle proposer periods, and then another six cents between the middle and late proposer periods.

Model 2 introduces a dummy for the  $2 \times 1$  treatment along with an interaction term between this dummy and the dummy for middle proposer periods. There is no need to include an interaction term for late proposer periods, since all of the observations in late proposer periods are from  $1 \times 1$  sessions. Neither of these two additional terms are statistically significant by themselves at any standard level, nor are they jointly significant ( $\chi^2 = 0.37$ ,  $p > .10$ , 2 d.f.). We detect no sign of a treatment effect for the proposers.

While we cannot detect a treatment effect for the proposers, we do observe substantial evidence that proposers' choices react to responders' actions. Model 3 modifies Model 1 by adding in the previously observed rejection rates for offers of 5, offers of 4, and offers of 3. The coefficients on all three rejection rates are positive and statistically significant at the

Table 3. Ordered probit regressions on proposer data (148 subject, 5012 observations).

Variable	Model 1	Model 2	Model 3
Middle proposer periods	.241** (.028)	.272** (.053)	.244** (.030)
Late proposer periods	.122** (.031)	.108* (.050)	.032 (.038)
2 × 1 Dummy		.036 (.048)	
Middle proposer periods × 2 × 1 dummy		-.051 (.063)	
Rejection rate (offer = 5)			5.792** (.574)
Rejection rate (offer = 4)			1.754** (.137)
Rejection rate (offer = 3)			1.227** (.075)
Log-likelihood	-3631.046	-3630.860	-3579.743

+ Significantly different from 0 at the 10% level.

\* Significantly different from 0 at the 5% level.

\*\* Significantly different from 0 at the 1% level.

1% level. The three rejection rates are also jointly significant at the 1% level ( $\chi^2 = 102.61$ ,  $p < .01$ , 3 d.f.). Not surprisingly, an increase in any of the rejection rates causes an increase in offers consistent with learning by proposers. Even though we are not finding the predicted treatment effect for proposers, the data contain clear evidence of learning by the proposers.

The lack of any observable treatment effect for the proposers is puzzling. While our primary interest is in the behavior of responders, the prediction of a treatment effect for proposers is just as robust as the prediction of a treatment effect for responders. Moreover, the results of Model 3 in Table 3 suggest that we are seeing exactly the sort of learning by proposers that the model predicts. The best explanation for this missing treatment effect derives from the forces underlying the reinforcement learning model's predictions.<sup>18</sup>

Through its manipulation of the relative speeds of play, the 2 × 1 treatment is supposed to give subjects different experiences with the distribution of offers (for responders) or the likelihood that a particular offer will be rejected (for proposers). The predicted treatment effects are driven by these expected differences in experience. If subjects receive different experience (on average) in the two treatments, they should learn to behave differently. If there were no observable differences between the two treatments in the experience subjects playing one of the roles receive, the reinforcement learning model would not predict a treatment effect. Given that a treatment effect is only observed for the responders, we hypothesize that it is possible from the point of view of a responder to observe differences between the behavior of proposers in the two treatments, but it is not possible from the point of view of a proposer to see differences between the behavior of responders in the two treatments.

Table 4. Regressions from subjects' perspective.

Ordered probit on offers from responders' perspective (102 subjects, 5058 observations)		Probit on responses from proposers' perspective (148 subjects, 5012 observations)	
Variable	Parameter estimate	Variable	Parameter estimate
Late periods	.143** (.030)	Constant	3.088** (.161)
2 × 1 dummy	-.178** (.051)	Offer	-1.013** (.022)
		Middle proposer periods	-.082 (.067)
		Late proposer periods	.111 (.077)
		2 × 1 dummy	-.068 (.110)
Log likelihood	-4813.007	Log likelihood	-1877.353

+Significantly different from 0 at the 10% level.

\*Significantly different from 0 at the 5% level.

\*\*Significantly different from 0 at the 1% level.

To test this hypothesis, we ran regressions looking at proposers' behavior from a responder's point of view and vice versa. The results of this analysis are shown in Table 4.

The left side of Table 4 reports an ordered probit regression run on offers from the responders' point of view. Unlike the regressions reported in Table 3, this regression uses periods rather than proposer periods so as to represent time from the point of view of a responder. Late periods are defined as in Table 2 to be periods later than the 15th period. A random effects specification is used to control for individual effects, where the individuals are the responders rather than the proposers. The goal is to run a regression using only information available to the responders.<sup>19</sup>

The central result of the regression on offers from a responder's point of view is that the dummy for the 2 × 1 treatment is negative and statistically significant at the 1% level. Looking at the data from a responder's viewpoint, we can find systematic differences between the distributions of offers in the two treatments. Since the experience (on average) that responders are receiving differs between the two treatments, the reinforcement learning model predicts that we should observe a treatment effect for responders. Indeed, a strong treatment effect is observed in the data.

The right side of Table 4 reports a probit regression on responders' behavior run from the proposers' perspective. Middle proposer periods and late proposer periods are defined as in the regressions reported on Table 3: middle proposer periods is a dummy for proposer periods after the 15th proposer periods and late proposer periods is a dummy for proposer periods after the 25th proposer period. A random effects specification is used to control for individual effects, with the individuals being proposers rather than responders. The regressions are designed to only use information that is available to the proposers.

The critical variable in interpreting this regression is once again the  $2 \times 1$  treatment dummy. While the coefficient on this variable has the correct sign, it is not statistically significant at any standard level. This result implies that the experience proposers are receiving in the two treatments does not differ significantly. As such, the reinforcement learning model does not predict that we will find a treatment effect, and indeed we don't.<sup>20</sup>

To summarize, the presence or absence of treatment effects for the two roles is consistent with the reinforcement learning model once we account for the presence or absence of observable differences in the experience subjects receive in the two treatments.

On a broad level, the experimental results are consistent with the reinforcement learning model. The reinforcement learning model predicts that responders adjust more slowly than proposers in the ultimatum game because their incentives to change their behavior are much lower. Because the relative speed of adjustment is slower for responders than for proposers, proposers stop making low offers before responders learn to accept them. The  $2 \times 1$  treatment is designed to manipulate the relative speeds of adjustment so that responders will both receive more low offers and receive them farther into the experiment. Having given the responders more relevant experience, we expect to see more adjustment towards accepting low offers. This is exactly what is observed. Rejection rates are lower in the  $2 \times 1$  treatment than in the  $1 \times 1$  treatment.

## 5. Conclusions

The experiment described in this paper is designed to elicit a subtle yet important effect. By successfully detecting learning by the responders, we show that a critical empirical prediction of reinforcement learning models is fulfilled. The data are consistent with the responders learning in exactly the same manner that the proposers learn. The famously anomalous behavior of responders in the ultimatum game need not depend on responders having unchanging preferences for fairness, but instead can be explained by the relatively slow speed of responders' learning as compared with proposers' learning.

Our point is not that the magnitude of learning by responders in the ultimatum game is especially large. The learning theories predict that responders will learn only slowly, and the positive evidence of a small effect collected here is consistent both with these predictions and with previous studies in which no responder learning was detected. It is because learning by responders is so difficult to detect that the responder learning detected here says as much about the learning models as does their success at predicting behavior in games in which learning is much more evident. Successful models should allow us to predict non-obvious effects that haven't yet been observed as well as the obvious ones we already know about. If learning models were only successful at predicting behavior in games in which rapid learning is evident, we could not be as confident that the models were capturing the cause of the learning behavior, and not just its gross effects. The successful observation of the predicted responder learning in the  $2 \times 1$  condition allows us to have greater confidence in the explanatory as well as the predictive power of learning theories.

Our results do not imply that theories of fairness have no role to play in understanding behavior in the ultimatum game and related games. The learning models do a good job of explaining how behavior evolves given a starting configuration of strategies, but do

not explain how this initial state arises. Theories of fairness have a natural role to play in explaining the initial state of play. Moreover, work on related games by Cooper and Stockman (2000) has shown that models that combine fairness and learning do a better job of capturing the major features of the data than models that use just fairness or just learning. We have done similar exercises with the data set reported in this paper (Cooper et al., 1999), and also find that a hybrid model combining fairness and learning outperforms a pure model. Thus, while our results indicate that learning by responders must be accounted for in understanding the ultimatum games, explanations of responder behavior in the ultimatum game also leave a role for theories incorporating preferences concerning fairness.

We are not the first experimenters to study learning by responders, but we believe our work represents a significant step forward in understanding learning by responders. In particular, we have found unambiguous evidence of learning by responders. Most experiments that allow players to gain experience reveal changes over time in the offers, but have been unable to detect changes in the acceptance/rejection behavior of responders. (See e.g. Roth et al., 1991; Slonim and Roth, 1998; Duffy and Feltovich, 1999).

The notable exception to this statement is List and Cherry (2000).<sup>21</sup> That paper replicates the earlier work of Slonim and Roth (1998) while adding an element of proposer entitlement.<sup>22</sup> In the low stakes treatment, no changes in responders' behavior are observed over the ten periods of the experiment. In the high stakes treatment, there is a statistically detectable decrease in the rejection rates by responders over the final three periods. List and Cherry do not attribute this change to any specific cause, but do note that it is consistent with the learning model of Roth and Erev.

Our work both complements the work of List and Cherry and adds to it substantially. List and Cherry note that their two differing treatments generate different distributions of offers, and conjecture that this difference in distributions drives the differing behavior by responders. Our results largely confirm this conjecture. Our work expands the work of List and Cherry in three specific ways, listed in order of increasing importance. (1) Because their purpose was not to study responder learning, List and Cherry's experimental design includes potential confounds that could explain their results. Our experiments are completely standard ultimatum games from the subjects' points of view with no potential confound. (2) The effect found by List and Cherry is an endgame effect, and is consistent with subjects abandoning a reputation for toughness. This suggests that the changes in responders' behavior they found may not necessarily be due to learning in the sense we typically think of it. The effects we find in our data are in no sense endgame effects. (3) We clearly demonstrate that changes in responders' behavior are driven by the stream of offers they are receiving. In other words, we aren't just seeing a decline in errors for some unknown reason, we aren't just seeing an endgame effect, but rather we are seeing true learning in which subjects are responding to their experiences by changing their behavior.

On a broader level, it bears repeating that while showing that responders learn is an important result, our purpose is more than this. The theoretical argument that this paper explores depends on differences in relative speeds of learning—the prediction that play in the ultimatum game need not converge to the subgame perfect equilibrium follows from the observation that while responders learn in the same fashion as proposers, they learn more slowly than proposers. This implies manipulating the relative speeds of learning can alter the behavior that we observe. Our goal, which we have achieved, was to verify this prediction.

## Appendix A

This appendix contains technical material describing the reinforcement learning model and explaining how simulations of this model were implemented.

### A.1. Description of the reinforcement learning model

The model for each individual is specified as follows. At time  $t = 1$  (before any experience has been acquired) each player  $n$  has an initial propensity to play his  $k$ th pure strategy, given by some number  $q_{nk}(1)$ . If player  $n$  plays his  $k$ th pure strategy at time  $t$  and receives a payoff of  $x$ , then the propensity to play strategy  $k$  is updated by setting  $q_{nk}(t+1) = \phi q_{nk}(t) + x$ , while for all other pure strategies  $j$ ,  $q_{nj}(t+1) = \phi q_{nj}(t)$ , where  $0 < \phi < 1$  is a “forgetting” parameter that regulates how slowly past experience decays. The probability  $p_{nk}(t)$  that player  $n$  plays his  $k$ th pure strategy at time  $t$  is  $p_{nk}(t) = q_{nk}(t) / \sum q_{nj}(t)$ , where the sum is over all of player  $n$ ’s pure strategies  $j$ . The model thus has two parameters,  $\phi$ , and the sum over all pure strategies  $j$  of a player’s initial propensities,  $S = \sum q_{nj}(1)$ . This latter parameter, which is taken to be the same for all players, is called the “strength” of the initial propensities, and influences the early speed of learning.

Thus this model predicts that strategies that have been played and have met with success tend over time to be played with greater frequency than those that have met with less success; i.e. these dynamics obey the “law of effect.” Also, the learning curve will be steeper in early periods and flatter later (because  $\sum q_{nj}(t)$  is an increasing function of  $t$ , so a payoff of  $x$  from playing pure strategy  $k$  at time  $t$  has a bigger effect on  $p_{nk}(t)$  when  $t$  is small than when  $t$  is large, i.e. the derivative of  $p_{nk}(t)$  with respect to a payoff of  $x$  is a decreasing function of  $t$ ).

### A.2. Implementation of the reinforcement learning model to predict the outcome of the experiment

In using the reinforcement learning model to predict the outcome of the current experiment, we begin by specifying a strategy set for each player. For the proposers, the strategy set equals the set of available offers,  $\{1, 2, \dots, 10\}$ . For responders, we follow Roth and Erev (p. 177) in limiting the set of available strategies to be cutoff strategies. A cutoff strategy specifies the lowest offer that a player would be willing to accept. For example, an individual with a cutoff of 4 will accept any offer greater than or equal to 4, and will reject lower offers. The set of available cutoffs corresponds to the set of available offers,  $\{1, 2, \dots, 10\}$ .<sup>23</sup>

Having limited responders to cutoff strategies, the learning model can be applied to the ultimatum game as a normal form game in which each of the two players chooses a number between 1 and 10 simultaneously. Let  $O$  be the offer selected and  $C$  be the cutoff selected. The following equations give the proposer’s payoff,  $\pi_P$ , and the responder’s payoff,  $\pi_R$ .

$$\begin{aligned} \pi_P(O, C) &= 10 - O \quad \text{if } O \geq C \\ &= 0 \quad \text{otherwise} \end{aligned} \tag{1a}$$

$$\begin{aligned}\pi_R(O, C) &= O \quad \text{if } O \geq C \\ &= 0 \quad \text{otherwise}\end{aligned}\tag{1b}$$

To predict differences between the  $1 \times 1$  treatment and the  $2 \times 1$  treatment, we ran 10,000 simulations of each treatment. The simulations of the  $1 \times 1$  treatment use two populations of ten individuals (represented by the learning model), ten proposers and ten responders, with all individuals playing in each period. The simulations of the  $2 \times 1$  treatment use three populations of ten players, two groups of ten proposers and ten responders. As in the actual experiments, the responders play each period while the two groups of proposers alternate periods. The proposers and responders playing in any particular period are randomly matched. All of the simulations last for 50 periods, the modal number of periods for the actual experiments. This allows us only 25 periods to compare proposers' behavior between the two treatments.

Our goal in running these simulations was to generate qualitative predictions for the effect of the  $2 \times 1$  treatment, not to find the best fit for some particular data set. Therefore, our approach is to choose a plausible set of parameters, determine the predicted treatment effects, and then see how robust these predictions are to changes in the parameters. For simplicity, we use the same initial probabilities over offers and cutoffs for all of the simulations and only vary the strength of initial propensities and the forgetting parameter. For simplicity, all individuals are assumed to have identical initial propensities.<sup>24</sup> The initial propensities for proposers put weights of 33.3% on an offer of 5, 33.3% on an offer of 4, 16.7% on an offer of 3 and 16.7% on an offer of 2. The initial propensities for responders put weights of 20% on a cutoff of 5, 40% on a cutoff of 4, 20% on a cutoff of 3, and 20% on a cutoff of 0. The implied initial reject rates are 0% for an offer of 5, 20% for an offer of 4, 60% for an offer of 3, and 80% for an offer of 2. Initially, an offer of 5 maximizes expected payoffs by a narrow margin. These initial propensities are in the ballpark of what is typically seen in ultimatum game experiments. In the baseline simulations, the initial strength of propensities was set equal to 10 for all individuals and the forgetting parameter was set equal to 1 for all individuals. We then vary these values to evaluate the sensitivity of the model's predictions to the parameters.

## Appendix B

This appendix contains technical material on the regressions described in Sections 4.1 and 4.3.

### *B.1. Regressions on responder data*

All of our econometric analysis of responders' data uses probit regressions. The dependent variable is always the responder's choice, with an acceptance coded as a 0 and a rejection coded as a 1. Thus, positive parameters correspond to a higher probability of rejection and negative parameters correspond to higher probability of acceptance. Statistical tests of

significance for individual parameter estimates are always two-tailed  $z$ -tests, and tests of joint significance are always log-likelihood ratio tests.

Many of our regressions include lagged variables. Due to this, the first period of data is deleted from our data set. Other than this we have used all observations from all responders.

Casual investigation of the data indicates that there are strong individual effects in the responder data. A random effects specification, allowing for correlation between observations from the same individual, is used to correct for these individual effects. The regression results strongly support the use of a random effects specification since the random effects term is always significant at the 1% level. We do not report the estimate of the random effects term in our tables, since it has no economic relevance.<sup>25</sup>

Several of our regressions include measures of past behavior by proposers. We use three variables to measure proposers' past behavior: the proportion of previous offers greater than or equal to 5 (high offers), the proportion of previous offers less than or equal to 3 (low offers), and the lagged offer from the preceding period. All three of these measures are calculated using the individual responder's past history—no information that a subject could not have observed is used. The first two variables allow us to control for the distribution of past offers without imposing linearity while the third controls for the possibility that the most recent experience gets extra weight. While these three measures of proposers' behavior are closely related to each other, they are far from being perfectly correlated.<sup>26</sup>

Since we are interested in learning effects, our analysis includes controls for changes in time. We use a non-linear specification for time, with the variable "Late Periods" being a dummy for observations after the 15th period. We tested a variety of alternative specifications for time, including linear specifications and non-linear specification with more intervals, and found that this one best fits the data. The choice of specifications for time does not affect any of our main conclusions.<sup>27</sup>

### *B.2. Regressions on proposer data*

Our econometric analysis of proposer data uses ordered probit regressions. Our use of this non-linear specification is driven by the discreteness of proposer data. Unlike many ultimatum game experiments in which subjects are choosing (approximately) over a continuum, proposers in our experiments only have a small number of possible strategies. We can therefore think of the observable choices as categories capturing subjects' underlying choices over the continuum of possible offers.

Offers are classified into three categories: offers less than or equal to 3, offers of 4, and offers of 5 or greater. The first period of data is deleted to allow for the use of lagged variables. Since there are strong individual effects in the proposer data, we use a random effects specification. The breakpoints between categories and the random effects term are not reported in our tables. While these items are always statistically significant, they lack any economic significance. The ordered probit regressions do not contain a constant since this would be colinear with the breakpoints.<sup>28</sup>

All regressions on proposer data measure time from the perspective of the proposer. Thus, the tenth period of a  $2 \times 1$  session is considered equivalent to the fifth period of a  $1 \times 1$  session, since in both cases the proposer is playing for the fifth time. We refer to the

“proposer period” to make it clear that we are referring to the number of times a proposer has played. We use proposer periods to avoid comparing apples with oranges. It isn’t surprising that a proposer playing for the tenth time is different from one playing for the fifth time, given the strong learning dynamic in the data. Instead, we are trying to find differences between subjects who have played the same number of times in different treatments.

As with responders, we use a non-linear specification for time with proposers. We subdivide proposer periods into three classes: proposer periods 1–15, proposer periods 16–25, and proposer periods 26–60. We refer to these as early, middle, and late proposer periods respectively. The break following 15 proposer periods is used to parallel our analysis of responders’ data. The break following 25 proposer periods isolates proposer periods that only contain data from  $1 \times 1$  sessions. This guarantees that differences we identify between the treatments aren’t due solely to the fact that there are twice as many proposer periods in the  $1 \times 1$  sessions.<sup>29</sup> The regressions include a dummy for proposer periods greater than 15 (“middle proposer periods”) and a dummy for proposer periods greater than 25 (“late proposer periods”). Thus, the parameter labeled “late proposer periods” reflects the difference between offers in the middle proposer periods and offers in the late proposer periods, not between offers in the early proposer periods and offers in the late proposer periods.

Model 3 on Table 3 modifies Model 1 by adding three measures of responders’ behavior. In choosing the variables to be included, many obvious measures of responders’ behavior could not be used because of concerns with endogeneity and sample censoring.<sup>30</sup> We employ the rejection rates for offers of 5, offers of 4, and for offers of 3. These three rejection rates are calculated at the session level. For example, the rejection rate for offers of 4 is based on the responses to all previous offers of 4 by any proposer in the same session as the subject whose behavior we are trying to predict. While the session rejection rate differs from the rejection rate observed by any one individual, session rejection rates should be highly correlated with the responses observed by individual proposers. Therefore we believe that the session rejection rates are good proxies for the responder behavior observed by an individual subject.<sup>31</sup>

### Appendix C: Instructions

The purpose of this session is to study how people make decisions in a particular situation.

During this session you are going to participate in several rounds of negotiations. In each round the group of people in the room is divided into pairs. Each pair will bargain on how to divide \$10.00.

#### *How do you bargain on the division?*

One of you, the PROPOSER, will propose a division of the money. The person who receives the offer, the RESPONDER, can either accept or reject it. In either case the round ENDS immediately after the response to the offer is made. If the RESPONDER accepts the proposed division, then the money is divided accordingly. If the RESPONDER rejects the proposed division the round ends in disagreement and EACH one of you receives nothing.

PROPOSER	RESPONDER
\$9.00	\$1.00
\$8.00	\$2.00
\$7.00	\$3.00
\$6.00	\$4.00
\$5.00	\$5.00
\$4.00	\$6.00
\$3.00	\$7.00
\$2.00	\$8.00
\$1.00	\$9.00
\$0.00	\$10.00

In this session, the PROPOSER can not demand more than \$9.00 for him/herself out of the \$10.00. Also, the amounts must be multiples of \$1.00. Hence, the PROPOSER can propose only one of the following 10 divisions:

YOU WILL HAVE THE SAME ROLE, PROPOSER OR RESPONDER, IN ALL ROUNDS. Your role will be determined randomly.

*How do you get paid for your participation?*

For completing the session you will receive \$5.00. During the session you will participate in a series of rounds described above. At the end of the session, *one of the rounds* will be chosen at random, and you will be paid in cash what you earned in that round, in addition to the \$5.00.

*Who are the bargainers?*

In each round the computer will randomly assign you to another person with whom you will bargain. You will never bargain with the same player twice in a row. Since you are interacting using the computer terminal, you will not know your co-bargainer's identity, nor will they know yours. These identities will not be revealed even after the session is completed.

*How to use the computer terminal*

The whole session is managed by a computer program. All exchanges (offers and responses) are done through the terminal in front of you. You key in offers/responses using the keyboard and watch for responses and other information by observing the screen.

The PROPOSER makes an offer by filling in his/her proposed share of the \$10.00 in the following statement on the screen: "I propose to get \$X.00 out of the \$10.00." The PROPOSER should type in one number to replace the X on the screen. After doing so the computer will display the proposal back and will ask for verification. At that point, the PROPOSER can revise his/her proposal if so desired.

A confirmed proposal will be sent to the RESPONDER who will be asked to accept or reject it, by typing A for accept and R for reject. After doing so the computer will ask for verification. At that point, the RESPONDER can change his/her decision if so desired.

At the bottom of the screen you will find a "History of Past Play" window. This window will display the outcomes of previous rounds you have played.

Note that a new round will start only after all games played in the previous round are finished. Since there are many players in this session, expect delays between rounds. Please be patient.

During the session you may not communicate with other participants expect through the computer terminals. If you have any questions during the session, raise your hand and a monitor will come speak to you.

Instructions will now be read out loud, and any questions will be answered.

### Acknowledgments

We would like to thank John Ham, Glenn Harrison, Hidehiko Ichimura, John List, and Bob Slonim for their helpful comments. As per usual, the authors of this paper are solely responsible for any errors it may contain.

### Notes

1. It was initially studied by Guth et al. (1982); see Roth (1995) for a survey of related work.
2. The term reinforcement learning is here meant broadly, and certainly includes the replicator dynamics used by Gale et al. (1995) to derive a similar explanation to the one discussed here.
3. In addition to the papers cited earlier, see also Bornstein et al., 1994, 1996; Cooper et al., 1997; Erev, 1998; Erev and Rapoport, 1998; Feltovich, 2000; Mookherjee and Sopher, 1994, 1997; Nagel and Tang, 1998; Rapoport et al., 1997, 1998; Roth et al., 2000; Sarin and Vahid, 2000; Tang, 1995; Van Huyck et al. 1994.
4. There also exists a literature on the evolutionary foundations for such preferences. See Samuelson, 2001, for a concise summary of this literature.
5. According to this view, the learning observed on the part of proposers is simply ordinary Bayesian learning about the parameters of responders' utility functions.
6. The reinforcement learning model also predicts a second order effect on proposer behavior, that reinforces the difference between the two conditions. The prediction is that, in the  $2 \times 1$  condition, since the (less experienced) proposers will be encountering responders who more quickly learn to accept low offers than in the  $1 \times 1$  condition, the proposers in the  $2 \times 1$  condition will learn not to make low offers more slowly than *equally experienced* proposers in the  $1 \times 1$  condition.
7. Average earnings were roughly \$9.00 for a two-hour session. Subject payments were made in cash immediately following the session.
8. A comparison of some of the newer learning models can be found in Erev et al. (2002).
9. See Erev and Roth (1998) for a discussion of variations on the basic reinforcement model. For an alternative approach to reinforcement learning that does not obey the Power Law of Practice, see Bush and Mosteller (1955) as well as Cross (1983).
10. Güth et al. (2001) find evidence that some individuals in ultimatum game experiments do not use cutoff strategies. However, given that 91% of their subjects *do* use cutoff strategies, we feel that the use of cutoff strategies is a defensible simplification. The model's qualitative predictions do not depend on this assumption.
11. Cooper et al. (1999) contains such a fitting exercise.
12. Specifically we set the initial strength equal to 10, and set the forgetting parameter equal to 1. Setting the rate of forgetting equal to 1 is equivalent to not including any forgetting in the model.

13. We consider all possible combinations of initial strength = 2.5, 10, and 40 and the forgetting parameter = .8 and 1.0. This gives us six sets of simulations including the baseline simulations. The choice of only six parameter combinations is made to keep the graphs relatively readable. Adding more parameter values would not affect our predictions. Likewise, we use the same initial probabilities for the players' strategies in all of the simulations to simplify matters, but our predictions would not change if we used a variety of differing initial probabilities.
14. Specifically, the strength of initial propensities is equal to 2.5 and the forgetting parameter is equal to .8.
15. All of our econometric analysis of responders' behavior uses the offer as an explanatory variable. To the extent that offers are endogenous, this could raise problems with our analysis. We are analyzing data at the individual level. There is negative correlation between the likelihood of receiving a low offer and the likelihood that responders in a particular session will reject such an offer. There must be positive correlation between the likelihood that a particular individual rejects a low offer and the likelihood that a responder randomly selected from his session rejects a low offer. Implicitly, we are assuming that our sessions are large enough that the product of these two correlations is small.
16. Beyond the main treatment variable, one unusual feature of our experiments is the large number of repetitions. One could argue that, compared with single-shot experiments, this might bias subjects towards using simple adaptive algorithms rather than trying to reason through the game. If this were true, we would expect first play behavior to be altered. However, comparing the first play behavior in our experiment with that observed in other ultimatum game experiments, we see no evidence to support this hypothesis. In particular, we gathered first period data from Güth et al.'s inexperienced sessions (1982), Forsythe et al.'s (1994) ten dollar ultimatum games, and Roth et al.'s (1991) ten dollar ultimatum games from Pittsburgh. The first two experiments are single-shot, with Forsythe et al. being run in the U.S. for the same stakes we used. The Roth et al. sessions repeated the game ten times, but used the same population and the same stakes as the current experiments. Comparing first play offers across these four experiments, our subjects are neither the most generous nor the least. To see this, we can break first offers into three categories: 50% of the pie or more, 40–49% of the pie, and less than or equal to 39% of the pie. For our experiment, the proportions of proposers in these categories are .439, .338, and .223 respectively. The respective proportions for the other experiments are .238, .167, and .595 for Güth et al., .750, .125, and .125 for Forsythe et al., and .593, .259, and .148 for Roth et al. Proposers from the Forsythe et al. experiments are the most generous while those from the Güth et al. experiments are the least generous. Looking at rejection rates, our subjects again look unexceptional. Breaking down the offers into the same three categories as above, we get rejection rates of .000, .243, and .355 for the current experiments. The respective numbers for the other experiments are .071, .143, and .280 for Güth et al., .000, .167, and .000 for Forsythe et al., and .188, .167, and .500 for Roth et al. These numbers should be taken with a grain of salt, since the number of observations in any category is relatively small. However, no clear ranking emerges among the four experiments for first period rejection rates.
17. We report marginal effects calculated at the average values of the independent variables. Given the non-linearity of the probit model, the predicted effect on rejection rates will differ for any particular offer.
18. Another possible explanation rests on the statistics being used. As indicated by the simulations, the treatment effect can be quite small while the individual effects in the data are substantial. Our ability to pick up treatment effects in the data depends on our ability to separate individual effects from systematic treatment effects. The more observations we have for each individual in our sample, the easier it becomes to disentangle individual and treatment effects. The experimental design gives us only 25 proposer periods in which we can make a direct comparison between the two treatments for proposers, as opposed to 50 periods for the responders. For all practical purposes, we are trying to detect a subtle treatment effect for proposers with a substantially smaller data set than the data set for responders. Not surprisingly, it is harder to detect a treatment effect with a smaller data set.
19. Unlike the regressions in Tables 2 and 3, neither regression in Table 4 includes an interaction effect between the treatment dummy and the time dummies. This simplified specification is used to maximize our chance of detecting a statistically significant treatment effect. Including interaction terms would not affect our conclusions.
20. It isn't surprising that we find a difference in responder's behavior between the two treatments in Table 2 but not in Table 4. The regressions in Table 2 use information that is not available to proposers, and therefore is not incorporated into the regressions in Table 4.

21. Harrison and McCabe (1996) also report an experiment on an ultimatum-like game where they find declining cutoffs for responders. Reflecting the primary purpose of their work, studying the effects of different varieties of feedback on subject behavior, the methodologies and the ultimatum game used in this experiment depart significantly from those typically used in ultimatum game experiments. Indeed, several of their sessions converge toward the perfect equilibrium, certainly not a standard ultimatum game result. The two sessions closest to standard ultimatum game experiments both in methods and results are their two control sessions U1 and U1' which show similar declines in the responders' cutoffs. This is interesting, because subjects are receiving no feedback whatsoever in U<sub>1</sub> while they receive standard feedback on their opponent's choice in U<sub>1</sub>'. This odd result, albeit based on a small sample, indicates that we need to be very careful about labeling changes in responder behavior as learning. Most models of learning, including all models of reinforcement learning, have players modifying their actions in response to feedback about previous outcomes. As such, establishing a link between the feedback and responders' choices is a necessary condition for a meaningful claim of learning. Changes by themselves may indicate nothing more than reduced subject confusion or greater introspection due to having more time to think.
22. Proposers were assigned to the high or low stakes payoff treatment based on their performance on a quiz. Responders were told that proposers had "earned an amount of money (the pie being bargained over) by participating in a previous session."
23. For symmetry with proposers (who cannot offer 0), we don't allow for the possibility that an offer of 10 will be rejected. Given the rarity of such offers, this is not an important restriction to the model.
24. Consistent with the use of a random effects specification in the econometric analysis of our data, we could have assumed that initial propensities are drawn from a random distribution. If we were trying to generate a best fit to the experimental data, such a specification could potentially play a valuable role. However, for our limited purposes this just adds extra complexity to the simulations without materially affecting the predictions.
25. More specifically, a random effects specification estimates the correlation between observations from the same individual ( $\rho$ ). This acts as an additional parameter in the maximum likelihood estimation. Rejecting the null hypothesis that this parameter equals zero tells us that observations from the same individual are not statistically independent. For example, for the key regression in Table 2, Model 2, the estimated value of  $\rho$  is .618 with a standard error of .013. This indicates a strong rejection of the null hypothesis of independence.
26. The correlation between the proportion of high offers and the proportion of low offers is  $-.550$ . The correlation between the proportion of high offers and the offer from the preceding period is  $.375$ . The correlation between the proportion of low offers and the offer from the preceding period is  $-.402$ .
27. To the extent that alternative specifications for time affect our results, we find that the treatment effect for responders widens significantly over time with some specifications. Since this result is not robust to the choice of specifications, we put little weight on it. We have also explored specifications in which the measures of past proposer behavior are interacted with the dummy for late periods. We do not find that these specifications significantly improve our ability to fit the data.
28. We limited the number of categories for offers to avoid convergence problems in the estimation. Given that over 90% of offers are in the 3–5 range, we lost little by eliminating separate categories for the extreme offers. As a robustness check, we have also run linear regressions with a random effect specification and ordered probits using all possible offers as categories and the method suggested by Moulton (1986) to correct the standard errors for clustering. In both cases, the qualitative results are identical to those reported in Table 3. This indicates that neither the use of an ordered probit specification nor the limited number of categories are driving our results.
29. Without adding in this extra time class, we find significantly lower offers in the  $2 \times 1$  sessions. This is a false positive, reflecting the fact that offers continue to rise beyond the 25th proposer period, the last proposer period in the  $2 \times 1$  sessions.
30. For a detailed discussion of the limitations we faced in constructing measures of responders' behavior, see the working version of this paper (Cooper et al., 1999).
31. We do not use the session rejection rates as instrumental variables *per se*. For our purposes there is little to be gained by using instrumental variables, and the technical problems raised by trying to use instrumental variables with an ordered probit regression would be substantial.

## References

- Blackburn, J.M. (1936). "Acquisition of Skill: An Analysis of Learning Curves." IHRB Report No. 73.
- Bolton, G. (1991). "A Comparative Model of Bargaining: Theory and Evidence." *American Economic Review*. 81, 1096–1136.
- Bolton, G.E. and Ockenfels, A. (1999). "ERC: A Theory of Equity, Reciprocity and Competition." *American Economic Review*, forthcoming.
- Bornstein, G., Erev, I., and Goren, H. (1994). Learning processes and reciprocity in intergroup conflicts. *Journal of Conflict Resolution*. 38, 690–707.
- Bornstein, G., Winter, E., and Goren, H. (1996). "Experimental Study of Repeated Team Games." *European Journal of Political Economy* Volume 12, Issue 4, 17-December-1996.
- Cooper, D.J., Garvin, S., and Kagel, J.H. (1997). "Adaptive Learning versus Equilibrium Refinements in an Entry Limit Pricing Game." *Economic Journal*. 107, 553–575.
- Cooper, D.J., Feltovich, N., Roth, A.E., and Zwick, R. (1999). "Relative versus Absolute Speed of Adjustment in Strategic Environments: Responder Behavior in Ultimatum Games." Working Paper.
- Cooper, D.J. and Stockman, C.K. (2002). "Fairness and Learning in a Step-Level Public Goods Game." *Games and Economic Behavior*. 41, 26–45.
- Duffy, J. and Feltovich, N. (1999). "Does Observation of Others Affect Learning in Strategic Environments? An Experimental Study." *International Journal of Game Theory*. 28, 131–152.
- Erev, I. (1998). "Signal Detection by Human Observers: A Cutoff Reinforcement Learning Model of Categorization Decisions Under Uncertainty." *Psychological Review*. 105, 280–298.
- Erev, I. and Rapoport, A. (1998). "Magic, Reinforcement Learning and Coordination in a Market Entry Game." *Games and Economic Behavior*. 23, 146–175.
- Erev, I. and Roth, A.E. (1998). "Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria." *American Economic Review*. 88(4), 848–881.
- Erev, I., Roth, A.E., Slonim, R.L., and Barron, G. "Combining a Theoretical Prediction with Experimental Evidence." Revised 2002.
- Fehr, E. and Schmidt, K.M. (1999). "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*. 114, 817–868.
- Feltovich, N. (2000). "Reinforcement-Based vs. Beliefs-Based Learning Models in Experimental Asymmetric-Information Games." *Econometrica*. 68, 605–641.
- Forsythe, R., Horowitz, J., Savin, N.E., and Sefton, M. (1994). "Fairness in Simple Bargaining Experiments." *Games and Economic Behavior*. 6, 347–369.
- Gale, J., Binmore, K., and Samuelson, L. (1995). "Learning to be Imperfect: The Ultimatum Game." *Games and Economic Behavior* 8, 56–90.
- Guth, W., Schmittberger, R., and Schwarz, B. (1982). "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior and Organization*. 3, 367–388.
- Harrison, G. and McCabe, K. (1996). "Expectations and Fairness in a Simple Bargaining Experiment." *International Journal of Game Theory*. 25, 303–327.
- List, J. and Cherry, T. (2000). "Learning to Accept in Ultimatum Games: Evidence from an Experimental Design the Generates Low Offers." *Experimental Economics*. 3, 11–30.
- Mookherjee, D. and Sopher, B. (1994). "Learning Behavior in an Experimental Matching Pennies Games." *Games and Economic Behavior*. 7, 62–91.
- Mookherji, D. and Sopher, B. (1997). "Learning and Decision Costs in Experimental Constant Sum Games." *Games and Economic Behavior*. 19(1), 97–132.
- Moulton, B. (1986). "Random Group Effects and the Prevision of Regression Estimates." *Journal of Econometrics*. 32, 385–397.
- 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.
- Ochs, J. and Roth, A.E. (1989). "An Experimental Study of Sequential Bargaining." *American Economic Review*. 79, 355–384.
- Rabin, M. (1993). "Incorporating Fairness into Game Theory and Economics." *American Economic Review*. 83, 1281–1302.

- Rapoport, A., Erev, I., Abraham, E.V., and Olson, D.E. (1997). "Randomization and Adaptive Learning in a Simplified Poker Game." *Organizational Behavior and Human Decision Processes*. 69, 31–49.
- Rapoport, A., Seale, D., Erev, I., and Sundali, J.A. (1998). "Coordination Success in Market Entry Games: Tests of Equilibrium and Adaptive Learning Models." *Management Science*. 44, 119–141.
- Roth, A.E. (1995). "Bargaining Experiments." *Handbook of Experimental Economics*, J. Kagel and A.E. Roth (eds.). Princeton University Press, pp. 253–348.
- Roth, A.E. and Erev, I. (1995). "Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term." *Games and Economic Behavior; Special Issue: Nobel Symposium*. 8, 164–212.
- Roth, A.E., Prasnikar, V., Okuno-Fujiwara, M., and Zamir, S. (1991). "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study." *American Economic Review*. 81, 1068–1095.
- Samuelson, L. (2001). "Introduction to the Evolution of Preferences." *Journal of Economic Theory*. 97, 225–230.
- Sarin, R. and Vahid, F. (2001). "Predicting How People Play Games: A Simple Dynamic Model of Choice." *Games and Economic Behavior*. 34, 104–122.
- Slonim, R. and Roth, A.E. (1998). "Learning in High Stakes Ultimatum Games: An Experiment in the Slovak Republic." *Econometrica*. 66(3), 569–596.
- Tang, F.-F. (1995). "Anticipatory Learning in Two-Person Games: An Experimental Study." Dissertation, University of Bonn.
- Thorndike, E.L. (1898). *Animal Intelligence: An Experimental Study of the Associative Processes in Animals*. Psychological Monographs, 2.
- Van Huyck, J.B., Cook, J.P., and Battalio, R.C. (1994). "Selection Dynamics, Asymptotic Stability, and Adaptive Behavior." *Journal of Political Economy*. 102, 975–1005.