Econometrica, Vol. 66, No. 3 (May, 1998), 569–596

LEARNING IN HIGH STAKES ULTIMATUM GAMES: AN EXPERIMENT IN THE SLOVAK REPUBLIC

BY ROBERT SLONIM AND ALVIN E. ROTH[1]

This paper reports an cxperiment involving an ultimatum bargaining game, played in the Slovak Republic. Financial stakes were varied hy a factor of 25, and behavior was observed both when players were incxpcricnccd and as they gained experience. Consistent with prior results, changes in stakes had only a small effect on play for incxpcricnccd players. But the present expcrimcntal design allows us to observe that rejections were less frequent the higher the stakes, and proposal\ in the high stakes conditions declined slowly as subjects gained experience. This Slovak experiment is the first to detect a lower frequency of rejection when stakes are higher and this can be explained by the added power due to multiple observations per subject in the experimental design. A model of learning suggests that the lower rejection frequency is the reason that the proposers in the higher stakes conditions of the ultimatum game learn to make lower offers.

KEYWORDS: Bargaining games, experimental design, learning.

1. INTRODUCTION

One of the conventions which has come to distinguish experimental economics from experimental psychology is that economics experiments typically attempt to control subjects' incentives by using monetary payoffs based on performance.[2] It is thus natural that one of the most frequent questions about experimental economics concerns whether hchavior observed when monetary incentives are relatively low can be generalized to similar environments with much higher risks and rewards. One way to address this is by within-cxpcrimcnt comparisons of behavior under widely different financial incentives, holding all else constant. The wider the range of payoffs the more powerful is the experiment at detecting potential differences in behavior that might he due to the size of the incentives. It is therefore attractive to conduct experiments in countries where the wage levels arc relatively low, so that subjects can be given large financial incentives with a given experimental budget.[3]

[1] This work was partially supported hy NSF Grant SES-9121968 to the University of Pittsburgh. We also thank Ido Erev, Nick Feltovich, Ellen Garbarino, Marjorie McElroy, and Jean-Francois Richard for helpful advice, and Alena Kimakova, Martin Mrva, and Gabriel Sipos for assistance in running the Slovak experiment. The current version of the paper reflects the contributions of several anonymous referees.

[2] See Roth (1995a) on the history of experimental economics, and the origin of monetary payments in economic\ experiments, starting with the critique hy W. Allen Wallis and Milton Friedman (1942) of the experiment reported by L. L. Thurstone (1931).

[3] A number of experiments have adopted this approach, e.g., in India (Binswanger (1980)), China (Kachelmeier and Shehata (1992)), Russia (Fehr and Tougareva (1995)), and Indonesia (Cameron (1995)). Another approach is to look for naturally occurring economic environments resembling

The present study reports an experiment conducted in the Slovak Republic in 1994, concerning how financial incentives influence observed behavior in an ultimatum bargaining game, a game that has an extreme perfect equilibrium that predicts that one side of the market will receive essentially none of the wealth. The stakes were varied by a factor of 25, from 60 Slovak Crowns (Sk) to 1500, with an intermediate stakes condition of 300 Sk. The smallest stakes condition (60 Sk) was chosen because it is similar to the experimental rewards per hour subjects get in experiments run in the U.S., where the stakes are often between 2 and 3 hours of wages. Subjects in the 60, 300, and 1500 sessions were bargaining over approximately 2.5, 12.5, and 62.5 hours of wages, respectively. The average monthly wage rate in the Slovak Republic at the time of the experiment was 5500 Sk.[4]

The ultimatum game consists of two players bargaining over an amount of money which we will call the "pie." One player, the proposer, proposes a division of the pie, and the second player, the responder, accepts or rejects it. If the responder accepts, each player earns the amount specified in the proposal, and if the responder rejects, each player earns zero. At perfect equilibrium the proposer receives all or almost all of the pie.

The ultimatum game has received a great deal of attention since the initial experiment by Guth, Schmittberger, and Schwartz (1982). It was studied, together with a related market game, under controlled conditions in a four country experiment by Roth, Prasnikar, Okuno-Fujiwara, and Zamir (1991). The game was played in ways that allowed the players to gain experience, and the play of the game revealed effects of experience; but behavior robustly showed no signs of approaching the perfect equilibrium. Furthermore, the observed transactions were most similar in the four subject pools when subjects were inexperienced, and became dissimilar in the different subject pools as subjects gained experience. Roth and Erev (1995) show that these observations are consistent with a simple model of learning. In the learning model, as in the experiment, small initial differences between subject pools become larger as subjects gain experience with the ultimatum game.

those whose robustness to higher stakes is of interest. This has the advantage of allowing truly high stakes to be examined, at the cost of losing the control available in the lab, and so allowing factors other than changes in stakes to influence the results. In this spirit, Telser (1995) identifies an aspect of salary negotiations in major league baseball as a high stakes analogue of the ultimatum game. Several investigators have looked to TV game shows for data. For example, Gertner (1993) studies risk attitudes on the game show *Card Sharks*, Metric (1995) investigates bidding behavior on the game show *Jeopardy*, and Berk, Hughson, and Vandezande (1996) study learning and bounded rationality on the game show *The Price is Right*.

[4] Statistics were unavailable on student wages. The 20 to 30 Sk per hour average student wage rate came from personal observation. In terms of purchasing power, for example, a dormitory room cost 150 Sk per month, a monthly bus pass cost 80 Sk, a local phone call cost 2 Sk for 6 minutes, and a movie cost 24 Sk. The exchange rate was 31 Sk for $1; thus the stakes were $1.9, $9.7, and $48.4 for the 60, 300, and 1500 sessions, respectively.

The design of the present experiment takes advantage of this observation to increase the power of the experiment to detect differences in behavior due to differences in stakes. Unlike previous high stakes experiments, the present experiment will give subjects an opportunity to play the game multiple times (with different partners) so that the effects of learning-which may magnify the effects of high stakes-can be observed.

Higher financial stakes might matter for several reasons. High stakes might reduce responders' willingness to 'punish' a given disproportionate offer, since it would raise the financial cost of indulging in such behavior. Likewise, high stakes might cause proposers to make proportionally less fair (smaller) offers to responders because higher stakes will raise the financial cost to make proportionally fairer offers. Also, proposers might make smaller proportional offers if they believe responders are more likely to accept a given disproportionate offer.[5] Hence, high stakes might move behavior towards the perfect equilibrium.

Controlled experiments reporting within-experiment comparisons of ultimatum games played for different stakes have generally found little effect on either offers or rejection frequencies. Roth et al. (1991) examined games played for $10 and for $30, and noticed no important difference. Straub and Murnighan (1995) also found little difference in proposer or responder behavior in ultimatum games between $5 and $100.[6] Hoffman, McCabe, and Smith (1996) found no significant difference in offers or rejection frequencies between $10 and $100 stakes in ultimatum games with either a random entitlement or contest treatment to determine the proposer. And Cameron (1995) found no difference in either proposer or responder behavior when stakes were changed from 5,000 to 200,000 Indonesian Rupiahs.

Except in Roth et al. (1991) (which considered only a modest variation in stakes), subjects in the experiments described above had no opportunity to obtain experience.[7] The results of Roth et al. suggest that the ultimatum game is a game in which experience serves to magnify initially small differences in behavior, and Roth and Erev (1995) present a learning model that predicts this. The current experiment therefore looks not only at a larger difference in stakes (a factor of 25) than has (with the exception of Cameron (1995)) previously been examined, but also looks at the effect of the difference as subjects gain experience. If the predictions of the learning model are correct, the interaction

[5] However, larger stakes may induce risk averse proposers to offer a greater share of the pie to avoid losing the greater monetary payoffs.

[6] Straub and Murnighan (1995) found, in their complete information condition, that the mean (median) lowest acceptable offer was constant at approximately 20% (15%) of the financial stakes level for pies of $10 to $100, in which subjects might get paid. The mean (median) lowest acceptable offer drops below 20% (15%) for stakes of $1,000 and $1,000,000 in hypothetical questions. The mean (median) offer was constant at approximately 40% (50%) for stakes between $5 and $80 and drops to about 35% (40%) for larger hypothetical stakes.

[7] Hoffman et al. (1996) investigated a one-shot environment in which subjects play one game, Straub and Murnighan (1995) obtained multiple offers and minimum acceptable offers from every subject, but subjects never received feedback from an opponent, and Cameron's (1995) subjects played two games, but with different stakes.

of stakes and experience should increase the power of the experiment to detect difference in behavior due to differences in the financial incentives.[8]

An additional advantage of having multiple (although nonindependent) observations per subject, even in the absence of learning, is that we are able to more precisely measure subtle differences in behavior caused by higher stakes. We lind the rejections were less frequent the higher the stakes, and proposals in the high stakes conditions decline as proposers gained experience. The ability to detect a significant difference in rejection frequency across stakes, which had eluded previous cxperimcnters. can be explained by the added power the current design provides. With the larger number of observations in the current design wc arc able to observe many slightly unequal proposals which are rejected only slightly less frequently when stakes arc higher, and we arc also able to observe a few very unequal proposals which arc rejected much less frequently when stakes are higher. And this difference in rejection frequencies, together with the opportunity which the experiment provides for proposers to learn from experience. allows us to detect differences in proposer behavior across stakes also.

The experimental design also includes sessions studying the market game examined by Roth ct al. (1991). The market game consists of players simultaneously making sealed bids for an indivisible object which has the same value to all players. The player who makes the highest bid earns the difference between the object's value and the highest bid, while all other bidders earn zero." The perfect equilibrium involves bidders bidding away all or almost all the wealth. Roth ct al. (1991) observed that behavior in the market game, unlike the ultimatum game. robustly and quickly converged to the perfect equilibrium as players gained cxpcricncc. Wc included the market game sessions because high stakes could have a different effect on behavior in the two games; in the market game high stakes give bidders more incentive to try to establish some implicit cooperation to keep bids down. Thus high stakes might cause behavior to move less towards perfect equilibrium in the market game and more towards perfect equilibrium in the ultimatum game. However, in the market game we could not detect any differences due to stakes: in all stakes conditions the transaction price quickly went to and remained at the perfect equilibrium. Because the results are very similar to those reported in Roth et al. (1991), the market game results will not be discussed in further detail.

[8] For some relevant review papers see Guth and Tietz (1990), and Roth (1995b). The paper by Fudenberg and Levine (1997) explores a nuanced approach to the issue of learning as a function of the costs of "irrational" behavior.

[9] See Roth et al. (1991) for a detailed description of the market game. The current market sessions differ from Roth et al. in that no experimental subject was assigned the role of seller. In Roth et al. one subject was the seller in each market, and could accept or reject the highest bid. (An active seller was used to control for fairness hypotheses in comparisons between the market and ultimatum games.) However, in all market games in all sessions, Roth et al. found the seller accepted every offer. Hence in this experiment all subjects are bidders, and the highest offer is automatically accepted. (The absence of an active seller reduces the set of Nash equilibria, but not the set of perfect equilibria.)

The paper is organized as follows: Section 2 describes the experimental design and equilibrium predictions for the ultimatum game, and Section 3 presents the experimental results, including a discussion of statistical power in different experimental designs. Section 4 briefly discusses how the results relate to learning behavior, and Section 5 concludes.

2. EXPERIMENTAL DESIGN AND PERFECT EQUILIBRIUM PREDICTIONS

In the ultimatum game, subjects participated in a sequence of ten games against different anonymous opponents.[10] During the ten game session a subject learned only the results of his or her own negotiations. Each subject was randomly assigned to be a proposer or responder, and a subject played the same role throughout the ten game session. In all games the pie 'was 1000 points and proposed divisions could be made in units of 5 points (0, 5, 10,. 995, 1000). The exchange rate for 1000 points was 60, 300, or 1500 Slovak Crowns (Sk), depending on the session. Ten ultimatum sessions were conducted, three at 60 Sk, four at 300 Sk, and three at 1500 Sk.

The subgame perfect assumption (with the additional assumption that subjects only want to maximize their monetary payoffs) means the responder will accept any positive offer, since rejecting any positive offer is inconsistent with wanting to maximize monetary reward. Since the smallest positive amount a proposer can offer is 5 points. no proposer will offer more than 5 points because responder will surely accept that amount. Thus, two subgamc pcrfcct equilibria exist: in one, proposer offers responder 5 points and keeps 995 for himself, and responder accepts (but would have rejected an offer of 0 points). In the other, proposer offers responder 0 points and responder accepts.[11]

3. EXPERIMENTAL RESULTS

A quick summary of our results is that, consistent with previous ultimatum game results (e.g., Straub and Murnigham (1995), Hoffmann et al. (1996), and Cameron (1995), we detect no significant difference between low and high stakes proposals or between low and high stakes rejection frequencies when examining inexperienced behavior (i.e., behavior in the first period). However, using all ten periods, we observe for the first time that responders in higher stakes reject proportionally equivalent offers less often, although rejections still occur even when substantial financial loss results. And when learning is examined, stakes also make a difference for proposals; offers decline in the higher stakes treatments as proposers gain experience. These results are described in more detail next.

[10] See Slonim (1995) for a complete description of the experimental design and procedures for the ultimatum sessions, which duplicate those described in Roth et al. (1991).

[11] In addition, in the ultimatum game any price can be observed at an imperfect Nash equilibrium.

TABLE I

SUMMARY OF ULTIMATUM GAME

Offers and Rejections by Range of Offers and Basic Statistics							
	60 Sk, N = 24		300 Sk, N = 33		1500 Sk, N = 25		
Offer Ranges	% Off	% Rej	% Off	% Rej	% Off	% Rej	#Offers
> 500	6.3 (15)	6.7 (1)	6.7 (22)	4.5 (1)	7.2 (18)	0.0 (0)	55
= 500	28 7 (69)	0.0 (0)	21.5 (71)	1.4 (1)	30.x (77)	1.3 (1)	217
450–495	21.7 (52)	9.6 (5)	22.7 (75)	5.3 (4)	6.0 (15)	0.0 (0)	142
400–445	24.6 (59)	23.7 (14)	21.8 (72)	12.5 (9)	32.4 (81)	4.9 (4)	212
350–395	11.3 (27)	40.7 (11)	9.4 (31)	9.7 (3)	5.2 (13)	0.0 (0)	71
300–345	4.6 (11)	45.5 (5)	10.6 (35)	22.9 (8)	7.2 (18)	11.1 (2)	64
250–295	2.5 (6)	66.7 (4)	3.9 (13)	30.8 (4)	3.2 (8)	37.5 (3)	27
< 250	0.4 (1)	100.0 (1)	3.3 (11)	90.9 (10)	8.0 (20)	60.0 (12)	32
All Offers	100.0 (240)	17.1 (41)	100.0 (330)	12.1 (40)	100.0 (250)	8.8 (22)	820
Offers < 500	35.1 (156)	25.6 (40)	71.2 (237)	16.0 (38)	61.6 (155)	13.6 (21)	54x
Average (all)	445		423		427		
Average (7 exclusions)	440		428		415		

Notes: The number in parentheses below each percent offer is the number of offers made in the range and the number in parentheses below percent rejected is the number of offers rejected in the range. The average (7 exclusions) removes all offers of the six subjects that made more than four offers greater than 50% and also excludes the one subject that made the offer of 5% of the pie in every round.

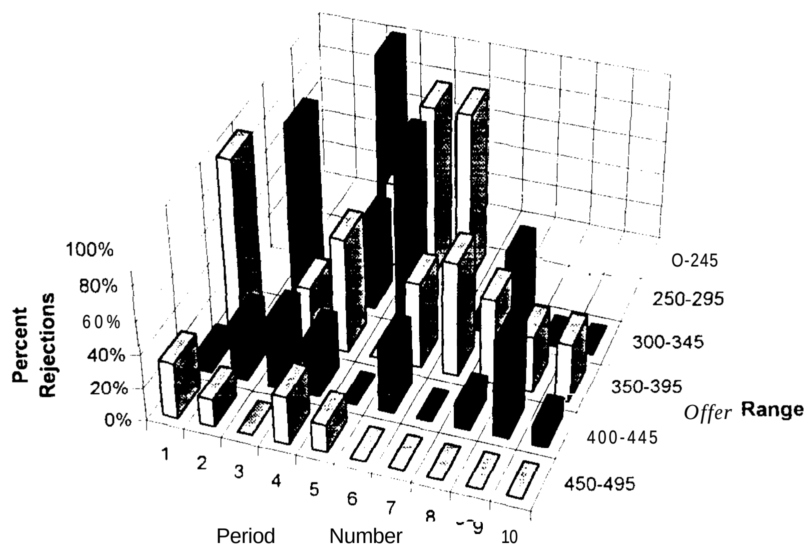
Table I describes proposer and responder behavior aggregating across rounds, and the Appendix provides a complete list of all players' choices. Table I can be read as follows; consider the offer range 400–445, which signifies proposer offered responder between 40 and 44.5% of the pie. In the 60 Sk condition, 24.6% (59/240) of all offers were in this range, and 23.7% (14/59) of these offers were rejected. Similarly. offers in this range accounted for 21.8% of the offers in the 300 Sk condition and 32.4% in the 1500 Sk condition, and these offers were rejected 12.5% and 4.9% of the time in the 300 and 1500 Sk conditions, respectively.

3.1 *Responder Behavior*

Overview: Over all offers, the rejection rate decreases from 17.1% (41/240) in the lowest stakes (60 Sk) to 12.1 % (40/330) and 8.8% (22/250) in the middle (300 Sk) and highest (1500 Sk) condition's, respectively. For disproportionate

offers, in which responders are offered less than half the pie, the rejection rate decreases from 25.6% (40/156) to 16.0% (38/237) to 13.6% (21/155) as the stakes increase.

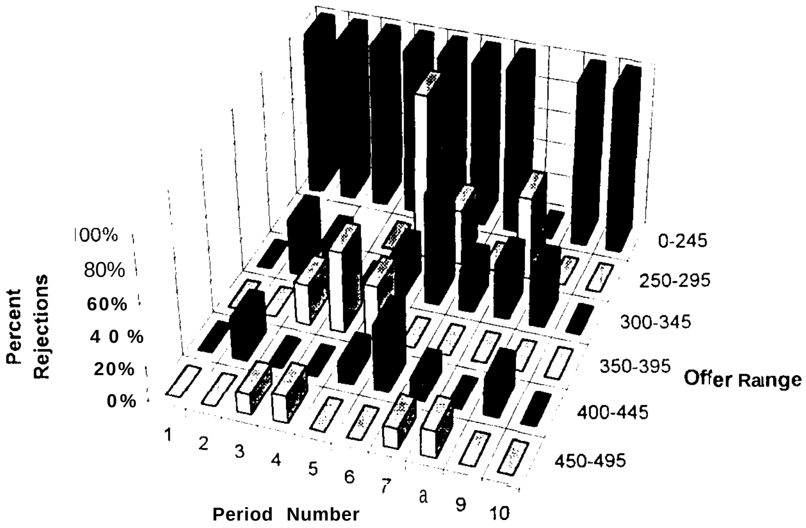
Figures la-lc show rejection rates over time by offer range. The height of each bar shows the percent of offers rejected for each period for a specific offer range. For example, in period nine 57% (4/7) of offers were rejected in the 60 Sk condition in the 400–445 offer range and in period ten 11% (1/9) were rejected. An empty square indicates no offers were made in that cell and a bar with no depth indicates offers were made but none were rejected. For example.



60 Sk: Rejections / Offers

	Offer Ranges						
Period	450 -495	400 -445	350 -395	300 -345	250 -295	0 -245	ALL
1	1/3	1/10	1/1	0/0	0/0	0/0	3/14
2	1/6	2/6	0/3	1/1	0/0	0/0	4/16
3	0/3	2/5	1/3	0/1	0/1	1/1	4/14
4	2/7	1/3	2/3	1/2	1/2	0/0	7/17
5	1/6	0/3	0/2	2/2	2/2	0/0	5/15
6	0/6	2/5	2/4	0/0	1/1	0/0	5/16
7	0/4	0/4	2/3	0/1	0/0	0/0	2/12
8	0/7	1/7	1/2	1/2	0/0	0/0	2/18
9	0/5	4/7	1/3	0/1	0/0	0/0	5/16
10	0/5	1/9	1/3	0/1	0/0	0/0	2/18
1-10	5/52	14/59	11/27	5/11	4/6	1/1	40/156

FIGURE 1a.—Low stakes (60 Sk).



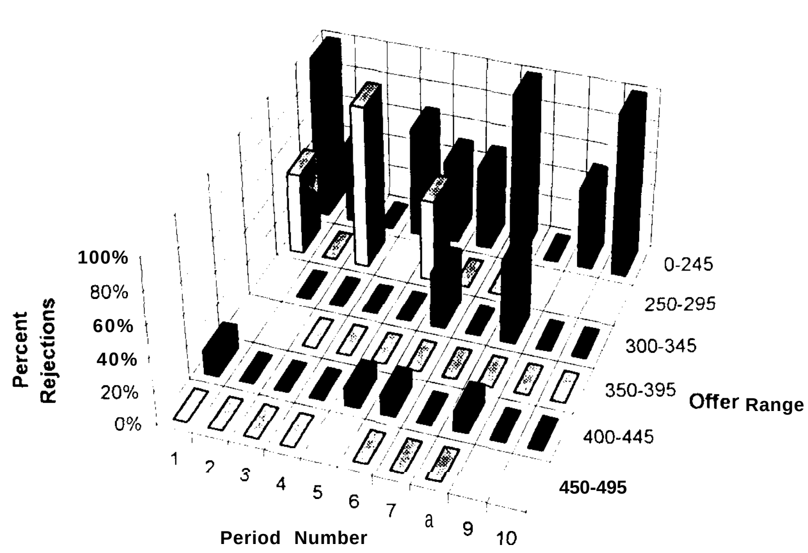
300 Sk: Rejections/Offers

	Offer Ranges 450 -495	400 -445	350 -395	300 -345	250 -295	0 -245	ALL
Period 1	0/11	0/3	0/2	0/3	0/0	1/1	1/20
2	0/11	1/4	0/3	1/3	0/0	1/1	3/22
3	1/8	0/6	1/4	1/4	0/0	1/1	4/23
4	1/6	0/8	1/2	0/4	0/1	1/1	3/22
5	0/8	1/8	1/3	1/4	2/2	1/1	6/26
6	0/7	4/10	0/1	2/3	1/3	1/1	8/25
7	1/8	1/8	0/3	1/4	0/1	2/2	5/26
8	1/6	0/8	0/4	1/3	1/2	0/1	3/24
9	0/5	2/9	0/6	1/3	0/2	1/1	4/26
10	0/5	0/8	0/3	0/4	0/2	1/1	1/23
ALL	4/75	9/72	3/31	8/35	4/13	10/11	38/237

FIGURE 1b.—Middle stakes (300 Sk).

in period ten of the 60 Sk condition no offers were made in the O-24.5 offer range whereas in the 450–495 offer range five offers were made hut none were rejected. Below each figure are the number of offers and rejections for each cell.

Figures 1a–1c highlight the main responder results that formal analysis will confirm. First, proportionally smaller offers arc rejected more often in all stakes conditions. Thus, in order to test the effect of stakes on rejections, it is important to control for the proportional size of offers. Second. the percent of offers t-ejected is smaller in higher stakes for each offer range less than 50% except in the 250–295 range. For example, for all ten periods in the 4X-405



1500 Sk: Rejections / Offers

	Offer Ranges 450 -495	400 -445	350 -395	300 -345	250 -295	0 -245	ALL
Period 1	0/4	1/7	0/0	0/0	1/2	2/2	4/15
2	0/4	0/6	0/0	0/1	0/1	1/2	1/14
3	0/2	0/8	0/2	0/1	1/1	0/1	1/15
4	0/1	0/7	0/2	0/1	0/0	2/3	2/14
5	0/0	1/10	0/1	0/1	1/2	1/2	3/16
6	0/1	1/9	0/1	1/3	0/1	1/2	3/17
7	0/2	0/8	0/2	0/1	0/1	3/3	3/17
8	0/1	1/9	0/1	1/2	0/0	0/2	2/15
9	0/0	0/8	0/3	0/4	0/0	1/2	1/17
10	0/0	0/9	0/1	0/4	0/0	1/1	1/15
ALL	0/15	4/81	0/13	2/18	3/8	12/20	21/155

FIGURE 1c.—High stakes (1500 Sk).

offer range, 9.6% (5/52) of offers arc rejected in the 60 Sk condition, whereas only 5.3% (4/75) arc rejected in the 300 Sk condition, and none (0/15) are rejected in the 1500 Sk condition. Third, offers are. in general, rejected fairly equally across periods for most offer ranges. For example, in the 300 Sk condition in the 450–495 offer range. no offers are rejected in the first two or last two periods and one offer is rejected in each of the third. fourth, seventh, and eighth periods.

To test responder behavior, we only investigate offers of less than 50%. For offers of 50% (or more). we predict (on the basis of earlier experiments) that virtually all offers will be accepted, regardless of pie size. and thus do not expect any difference due to stakes.[12] For offers less than SO%, responders may obtain utility not only from monetary payoffs, but also from punishing an unfair offer." Higher stakes may decrease rejections if the monetary reward dominates punishment value at higher stakes while punishment value dominates the monetary reward at lower stakes. (However. stakes may not have this effect if, as stakes increase, a responder's utility from punishing a proportionally small offer rises at least as much as his utility from money increases.)

First Round Behavior: A number of previous studies of ultimatum games compare aggregate rejection rates for different stakes. In the present experiment, for all disproportionate offers made in the first round, 21% (3/14), 5% (l /20), and 27% (4/15) were rejected by low, middle, and high stakes responders, respectively. None of the pairwise differences are significant." This result is similar to previous ultimatum game results discussed above. One concern with this result is the power to detect differences due to sample size; recall, there are 24. 33, and 25 responders in the three conditions and only 59.8% (49/82) of the offers in the first period are less than 50%.[15] A second concern is that differences in proportions offered between conditions are ignored. For example, there are no offers less than 30% in the lowest stakes in the first round, whereas there are five offers less than 30% in the middle and high stakes, and 4 of these offers are rejected, constituting all hut one of the rejections by middle and high stakes responders. Thus, looking at overall rejection rates may hide differences that exist among proportionally similar offers.

To control for proportionally equivalent offers, the following logit models were investigated for first period rejection behavior:

(1) $Reject = f(a + b_{o,,} * off)$,

(2) $Reject = f(a + b_{o,f,} * off' + h_{,,,} * pieM + b_{n} * pieH)$.

where *Reject* equals I if the offer is rcjcctcd and equals 0 otherwise, $f(x) = 1/(1 + e^{\cdot})$ is the logit function, *off* is the proportion of the pie offered (from 0 IO 49.5%), *pieM* = I if stakes arc 300 Sk and 0 otherwise (which measures the

[12] Table I chows that for offers greater than or equal to 50%, the proportion of offers (about 1/3) and the number of offers rcjcctcd (1 or 21 are nearly identical across stakes.

[13] See, for example, Bolton (1991) and Bolton and Zwick (1995).

[14] Two-tailed test of proportion results arc: low vs. middle: ı = 1.46, p = ,143: low vs. high: $z = -0.33$, $p > .70$; middle vs. high: $z = -181$, $p = .070$. Note. the middle stakes responders rcjectcd less often than the high stakes responders, counter to the expected direction.

"Hoffman. McCabe, and Smith (1996) had a similar sample size (24 and 27 subjects in $10 and $100 conditions) and similar results for a one shot game with random entitlement: 12.5% (3/24) and 18.5% (5/27) of offers were rejected in their low and high stakes, respectively.

marginal change in rejections from the lowest to middle stakes) and *pieH* = 1 if stakes are 1500 and 0 otherwise. The first model tests whether the proportion offered influences the probability of an offer being rejected, restricting the effect of stakes to have the same influence on rejections. Model 2 tests whether stakes influence rejections, controlling for the proportional offer.

Table II reports logit regression results. Columns 1 and 2 report the results for models 1 and 2. respectively. Subgame perfection predicts all positive offers will be accepted: thus the null hypothesis is $b_{off} = 0$. If smaller proportional offers are rejected more often, then $b_{off} < 0$ (i.e., larger proportional offers are rejected less often). In both models, b_{off} is significantly less than 0, indicating smaller offers are more likely to be rejected (models 1 and 2, $p < .01$).

Model 2 tests the effect of stakes on rejections. If stakes have no influence on rejections, then b_m = h,, = 0. If higher stakes reduce the likelihood that an offer will be rejected, then h,, < h,,, < 0. Model 2 results indicate that middle stakes responders are least likely to reject an offer and lowest stakes responders are most likely ($b_m = -4.61 < b_h = -1.17 < 0$). Although, high and middle stakes responders are directionally less likely to reject offers than low stakes respon-

TABLE II

LOGIT REGRESSION RESULTS: PROBABILITY OFFER IS REJECTED

	Round 1		All Rounds			
Parameter	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Intercept	4.22	7.08*	2.93***	4.29***	4.66***	4.39***
b_{off}	−15.7**	−20.3**	− 15.8"'	−17.6***	17.5***	17.7***
b_m		−4.61 (p = .13)		− 0.73' (p = .028)	− 0.69' (p = .037)	− 0.78' (p = .023)
b_h		− 1.17 (p = .35)		1.30** (p = .002)	1.29** (p = .002)	−1.39** (p = .001)
b_{acrej}			5.54***	5.29***	5.30***	5.49***
h_{round}					−0.07 (p = .156)	
$b_2, \ldots,$ h,,,						1^x
# Observations	49	49	54x	548	548	548
− 2 Log Likelihood	30.08	23.95	336.28	325.15	323.12	311.04
Model Comparisons:		vs. model 1 $\chi^2_{(2)}$ = 6.13 (p = .046)		vs. model 3 $\chi^2_{(2)}$ = 11.13 (p = .0038)	vs. model 4 $\chi^2_{(1)}$ = 2.03 (p = .154)	vs. model 4 $\chi^2_{(2)}$ = 14.1 ($p < .118$)

Notes 1^x − parameter estimates for round dummy variables not shown *p < 05, **p < 01, ***p < 001 ders, neither condition alone is significantly different from the low stakes condition (middle stakes, $p = .13$; high stakes, $p = .35$).[16]

In summary, we cannot reject that increasing stakes has no effect on the rejection rate in the first round. However, by looking at behavior across rounds, we can more powerfully investigate behavior for proportionally similar offers.

Behavior Across Rounds: In offer ranges less than 50% shown in Table I and Figures Ia-Ic, the rejection rate monotonically decreases as the financial stakes increase in every range except the 250–295 range. For example, in the 350-395 range, 40.7% (11/27), 9.7% (3/31), and 0% (0/13) of offers are rejected in the low, middle, and high conditions. In each of the four ranges in which there are at least 10 offers in each treatment, the rejection rate is always lower in the higher stakes conditions.

To test if rejections decrease as stakes increase, the following logit regressions were run:

$$(3) \qquad Reject = f(a + b_{off} * off + b_{avrej} * avrej_i),$$

$$(4) \qquad Reject = f(a + b_{off} * off + b_m * pieM + b_h * pieH + b_{avrej} * avrej_i),$$

where off, $pieM$, and $pieH$ are defined above. $Avrej_i$ equals the average number of offers rejected by subject i, excluding the current offer.[17] $Avrej_i$ is included to capture individual rejection propensity differences, since multiple observations of the same individual are not independent.[18] We expect $b_{avrej} > 0$; the more often subjects reject other offers, the more often they will reject the current offer.

Column 3 and 4 of Table II report the results. Model 3 and 4 results indicate that larger proportional offers decrease the likelihood that an offer will be rejected ($b_{off} < 0, p < .001$) and the more often responders reject other offers, the more often they will reject the current offer ($b_{avrej} > 0$, $p < .001$). Model 4 tests the influence of stakes on rejections. The results indicate that both the middle and high stakes conditions decrease the likelihood that an offer will be rejected relative to the lowest stakes condition (b_m = -0.73, $p = .0280$; $b_h = -1.30$, $p = .0016$).[19]

Figure 2 graphs the effect of stakes on rejections by proportional offer as predicted by model 4.[20] To compare the predicted to observed behavior, the graph includes actual rejection rates for each offer range reported in Table I. The model predicts that the higher the stakes, the less likely an offer will be rejected. The graph shows that the largest absolute difference between stakes in the likelihood to reject occurs for moderately disproportionate offers and that the smallest absolute difference occurs for offers very close to an equal split and for extremely disproportionate offers. For example, an offer of 45% (close to an equal split) is predicted to be rejected 9.4% of the time by low stakes responders and 1.5% of the time by high stakes responders. Similarly, an offer of 5% (an extremely disproportionate offer) is predicted to be rejected 99.2% of the time by low stakes responders and 94.4% of the time by high stakes responders. The absolute difference is much wider for moderately disproportionate offers; for example, an offer of 25% is predicted to be rejected 77.8% of the time by low stakes responders but only 33.4% of the time by high stakes responders.

To test whether rejection rates changed over time, we investigate two specifications:

$$(5) \qquad Reject = f(a + b_{off} * off + b_m * pieM + b_h * pieH + b_{avrej} * avrej_i + b_{round} * round),$$

$$(6) \qquad Reject = f(a + b_{off} * off + b_m * pieM + b_h * pieH + b_{avrej} * avrej_i + b_1 * r1 + \cdots + b_9 * r9)$$

Model 5 investigates whether rejections increase or decrease over time by including the variable *round*; round equals 1 for round 1, equals 2 for round 2,

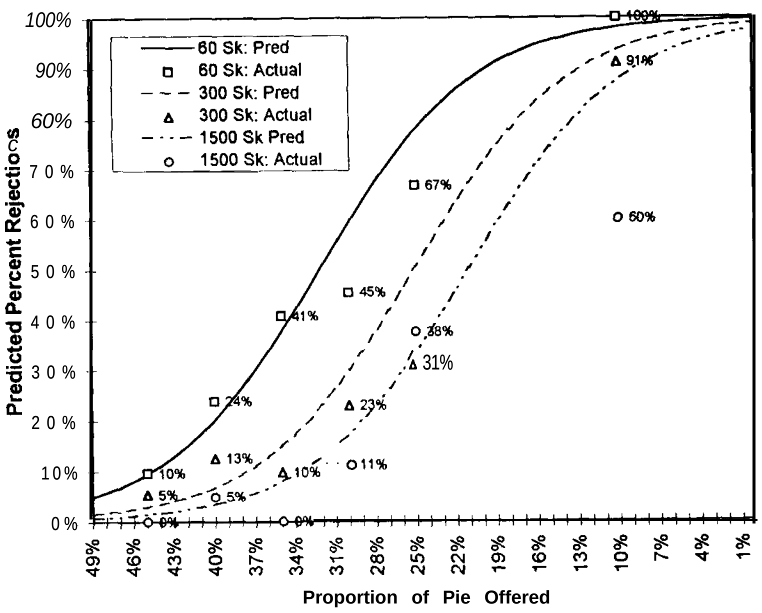
[16] The model 2 χ^2 test result indicates that compared to the restricted model 1 with $b_m = b_h = 0$, the likelihood that an offer will be rejected is significantly different across the three stakes conditions ($p = .046$). However, since model 2 parameter estimates indicate that middle stakes responders are less likely than high stakes responders to reject an offer, we cannot conclude that higher stakes cause offers to be rejected more often. Combining the middle and high stakes (i.e., restricting $b_m = b_h$), but otherwise using a model identical to model 2, higher stakes marginally decrease the likelihood of an offer being rejected ($p = .09$). However, we have no a priori reason to combine these two conditions and combining the lower two stakes conditions (i.e., restricting $b_m = 0$), but otherwise using a model identical to model 2, higher stakes (insignificantly) increase the likelihood of an offer being rejected ($p = .43$). In other words, middle stakes responders are less likely than either low or high stakes responders to reject an offer in period 1. Thus, depending on how we aggregate the three stakes conditions, we may draw different conclusions. When we analyze all ten rounds, this concern disappears. The limited number of disproportionate offers in period 1 stresses the importance of the low power to detect differences. This low power using just one period will be demonstrated below.

[17] For example, responder 21 received offers less than 500 in rounds 2, 4, 5, 6, and 8 and rejected offers in rounds 4 and 5. $Avrej_{21}$, thus equals .50 (2/4) in rounds 2, 6, and 8 and equals .25 (1/4) in rounds 4 and 5.

[18] Since 24, 33, and 25 subjects are in the three respective stakes conditions, the sample size is too small to use a random effects model to control for subject effects. Since subjects are nested within a single stakes condition, and further, since 38% (9/24), 52% (17/33), and 56% (14/25) of the subjects in the respective stakes conditions never reject an offer, a fixed effects model to control for subject effects is inappropriate (i.e., there is no variance for subjects who never reject). The variable $avrej_i$ is thus used as a proxy to control for subject effects.

[19] We also tested whether the effect of offers on rejections depends on the stakes condition by including in model 4 the interaction terms offer by $pieM$ and offer by $pieH$. The results of this test were that neither interaction term had any influence on rejections ($p > .90$ for both interaction terms), indicating that the effect of offers on rejections is independent of the stakes condition (and that the effect of stakes on rejections is independent of the offer).

[20] Figure 2 assumes the average rejection rate ($avrej_i$) for a hypothetical responder is at the mean of all experimental responders for each condition: 25.6%, 16.0%, and 13.0% in the low, middle, and high stakes conditions, respectively (see Table I, offers < 500).



Actual Rejection Rates:

	Offer Ranges: 450-495	400-445	350-395	300-345	250-295	0-245
Pie Sizes 60	9.62%	23.73%	40.74%	45.45%	66.67%	100.00%
300	5.33%	12.50%	9.68%	22.86%	30.77%	90.91%
1500	0.00%	4.94%	0.00%	11.11%	37.50%	60.00%

FIGURE 2. Rejection predictions (from regession model 4).

and so on. Round captures monotonic trends in rejection rates over time.[21] Model 6 includes dummy variables for each round to investigate whether rcjcction rates depend on particular rounds (for example, the first or last), possibly nonmonotonically. The results of both specifications indicate that rounds have no cffcct on rejection rates. In model 5, proportionally equivalent offers arc less likely to bc rejected over time ($b_{round} = -0.07$), but not significantly ($p = .16$). In model 6, round dummy variables do not significantly increase the explanatory power of the model ($\chi^2_{(9)} = 14.1$, $p = .12$). Two individual rounds

[21] The Roth-Erev learning model predicts that rejections will slowly decrease over time; thus model 5 can be viewed as a test of whether experimental responders also reject less often over time. However, in the Roth-Erev model, rejection rates fell very slightly over ten rounds and thus it may be difficult to detect this small decrease.

were significantly different than all other rounds; rejections were marginally higher in the 6th round ($p = .062$) and significantly lower in the tenth round ($p = .019$).[22] We interpret 6th round behavior as likely due to noise. The significantly lower rejection rate in the last round may signify an end effect or may also be noise. Thus. round has no systematic effect on rejections over time.

Statistical power: One question that naturally arises from the preceding analysis is why no significant differences in rejection frequencies are detected between stakes in the first period (or in one-shot experiments) whereas across all ten rounds we detect significantly fewer rejections in the higher stakes. One hypothesis is that there was an interaction effect in which rejection rates decreased over time in higher stakes more than in the low stakes. We tested this hypothesis by including the interaction of round by middle stakes and round by high stakes in model 5. However, neither interaction term has any effect on rejections ($p > .90$ for both interactions), indicating that the effect of round on rejections is the same across stakes conditions; i.e., the relative difference in the frequency of rejections between stakes is constant across rounds.[23]

Since stakes have an overall effect on rejections, but the difference is not observed in the first period nor is it observed to change over time, the inability to detect a significant difference in the first period (or in one shot experiments) may be due to low power.[24] The low power is likely caused by the fact that only small differences in responder behavior occur for offers near an equal split (recall Figure 2 and that the absolute difference between low and high stakes responders rejecting an offer of 45% is less than 10%) combined with the observation that the majority of offers are near the equal split (Table I reports that over 75% (626/820) of all offers are at least 40%). Thus, detecting a difference in responder behavior requires many observations to detect the small differences for nearly equal offers or to generate enough very unequal offers for which the difference in responder behavior is large.

To investigate the power to detect a significant difference, we generated 500 simulated data sets based on the model 4 results in which high stakes responders arc less likely to reject proportionally equivalent offers than low stakes respon-

[22] To test whether a round was distinct from all other rounds, ten separate regressions were run, each time including only one dummy variable for each round.

[23] We also ran models 1 and 2 for tenth period behavior in order to test whether stakes had a significant cffcct on rejection frequencies that may have developed after ten periods. However, no substantive differences between the model results for the first period behavior or tenth period behavior were observed; in both the first and tenth period lower offers significantly cause higher rejection frequencies and stakes have no significant cffcct on rejections. Thus, the effect of stakes on rejections appears to be constant across rounds.

[24] For example, Hoffman et al. had 24 and 27 responders in their one shot random entitlement ultimatum game, nearly identical in size to our 24, 33, and 25 responders in the low, middle, and high stakes conditions--and they observed 12% (3/24) and 18.5% (5/27) rejections in their low and high conditions, also similar to the 21%, 5%, and 27% we observed in the low to high conditions.

TABLE III

POWER TEST RESULTS

p-values	Round 1			All Rounds			
	b_m	b_h	b_{offer}	b_m	b_h	b_{arrej}	b_{offer}
$p < .10$	15%	15%	92%	97%	100%	100%	100%
$p < .05$	3%	2%	84%	90%	100%	100%	100%
$p < .01$	0%	0%	20%	52%	99%	100%	100%

Notes: The percent listed in each cell represents the power to detect a significant difference for the parameter estimate listed in the column header at the α significance level listed for the row. For example, the power to detect that $b_m \neq 0$ at the 10% significance level for the 1 Period Slovak Sample Size is 15%. In other words, if the identical experiment is run again, then there is a 15% likelihood that we would detect a difference between the rejection rate of the middle and low stakes responders in the first period at the 5% significance level.

ders. We then analyzed each data set identically to the analysis presented above. To generate the simulated data sets, simulated offers are set equal to the actual Slovak offers. Responder decisions are based on the behavior predicted by model 4; given an offer in the specific stakes treatment, model 4 is used to determine the *probability* that the offer is rejected: then a random draw is used to determine if the offer is rejected.[25] Table III presents the results of the analysis for the 500 data sets.

The first three columns of Table III indicate how often, using only first period data, we can detect the (known) difference between stakes generated from model 4. The power is extremely low: the power to detect a difference at even the generous 10% significance level between the low and middle or the low and high stakes is only 15%. The power to detect differences at the 5% significance level is less than 5%. In other words, if the experiment is repeated many times, we would expect to detect the known difference less than one time in twenty at the 5% level. In contrast, the power to detect that offers affect rejections at the 5% level is 84%. In other words, the sample size is sufficient to detect the substantial effect of offers on rejections using only first period data, but is not large enough to detect the more subtle effect of stakes on rejections. Thus, it is

[25] To determine the rejection probability for a given offer and stakes condition, model 4 also requires an average rejection rate ($arrej_i$) for each responder. We thus gave each simulated responder in a given stakes condition one of the Slovak responder's average rejection rates for that condition. For example, since there were 9, 17, and 14 responders in the low, middle, and high stakes who never rejected an offer, we include 9, 17, and 14 simulated responders in the low, middle, and high stakes who have an average rejection rate of 0%. Finally, we matched simulated responders to simulated proposers identically to how Slovak responders and proposers were matched. These procedures substantially reduce noise in the power tests and avoid making additional distributional assumptions about the determination of both offers and responder rejection propensities. Slonim (1997) investigated the power tests for a variety of distributional assumptions for both proposers and responders and found similar results to those presented below.

not surprising that we (and prior experiments using similar sample sizes) are unable to detect differences in rejection frequencies in the first period.[26]

The last four columns of Table III report power test results when using all ten periods. The power to detect a difference at the 5% level between the low and middle stakes is now extremely high (90% power) and at the 5% level we always detect the difference between the low and high stakes (100% power).

In summary, higher stakes responders are more likely to behave consistently with subgame perfect equilibrium in the sense that they reject fewer offers for proportionally equivalent shares of the pie. These effects are most significant when stakes differ by a factor of 25 and are also significant when the stakes differ by a factor of 5. Comparing these results with first round results and results from previous studies (which do not detect differences in responder behavior) indicates the value of multiple observations per subject; in first round behavior and one-shot games significant differences are not detected.

Though responders were generally more willing to accept proportionally smaller offers in higher stakes, it was not the case that proposers could make small offers with impunity; some responders rejected substantial monetary sums. For example, three out of 22 responders rejected a 40% offer in the high stakes condition one time, thus sacrificing 600 Sk (20 to 30 hours wages). Further, Y out of 16 offers between 20 and 24.5% (300 to 370 Sk) were rejected. Hence, higher stakes decreased the willingness of responders to reject disproportionate offers, but did not cause behavior to be consistent with perfect equilibria even when it cost one or more days' wages.

3.2. *Proposer Behavior*

Higher stakes may induce proposers to make lower offers for at least two reasons. First, proposers may obtain utility from both monetary rewards and fairness (Ochs and Roth (1989), Bolton (1991)); at lower stakes fairness may outweigh monetary rewards but at higher stakes monetary rewards may outweigh fairness. Second, if as observed, rejections decrease as stakes increase, expected payoffs may be maximized at lower offers. (If proposers are risk averse, this latter implication may not hold.)

To investigate the effect of stakes on offers, we do not analyze the small group of subjects who made a substantial number of offers greater than 50%

[26] Prescriptively, we investigated what sample size is needed to achieve adequate power so that we could confidently expect to observe the difference in stakes predicted by model 4. We increased the sample size and found that not until a sample size of 5 times the Slovak sample size are we able to achieve at least 75% power to detect a significant difference at the 5% level between the low and middle stakes conditions and not until a sample size of 4 times the Slovak sample size are we able to detect a significant difference at the 5% level between the low and high stakes conditions. Thus, increasing the stakes by a factor of 25, one would need approximately 100 responders in each condition to detect a significant stakes effect at the 5% level with 75% power when analyzing first period behavior (or a one shot experiment).

since we do not study (nor propose a model for) this particular behavior." The data, after removing subjects who made at least four offers greater than 50%, contain no subject who made more than 2 offers above 50%. Note that offers greater than 50% occurred almost equally in each stakes condition (about 7%) and in each round: thus removing them does not systematically influence a particular round or stakes condition. We also exclude subject number 401 from the analysis. This subject's offer in all ten rounds was 5 (5% of the pie), which was rejected in all but the eighth round.[28] We exclude this subject because his average offer was 3 standard deviations below the next lowest subject's average offer (220 by subject number 1003) and 5 standard deviations below the average offer of all subjects average offers. The exclusion of this subject has no significant effect on the results. After removing subjects who made more than two offers greater than 50% and one who always offered .5%, there are 23, 29, and 23 subjects in the low, middle, and high conditions, respectively.

Comparing first round offers across stakes, mean (median) offers are 451 (465), 460 (480), and 423 (450) in the low, middle, and high stakes conditions. Although offers are lower in the highest stakes condition, pairwise comparisons cannot reject that offers are the same across stakes (one-tailed t tests and Wilcoxian, Median, and Kolmogorov–Smirnov nonparametric tests cannot reject no difference; $p > .05$ for every pairwise comparison). This inability to reject that stakes do not influence offers is consistent with the results of Hoffman et al. (1996) and Cameron (1995).

The current design gives us the opportunity to test whether having multiple observations per subject may enable us to detect any significant differences. Figure 3a shows average offers over time. Notice that middle and low stakes average offers are similar in the first two rounds and both higher than high stakes offers, but for the last six rounds middle and high stakes average offers are similar and both lower than low stakes offers. The middle stakes offers tend to decrease the most over time, while low stakes offers tend to neither increase nor decrease consistently over all ten rounds.

Using offers across all rounds, the following analysis of variance was run:

$$(7) \qquad OFFER = PIE + ROUND + SUB(PIE) + PIE * ROUND,$$

[27] We removed proposers who made at least four offers greater than 50%. The result was that 1, 3, and 2 proposers were removed from the analysis in the low, middle, and high stakes conditions. All remaining proposers made no more than two offers greater than 50%. The number of offers above 50% that were removed was almost exactly 2/3 of the total number of offers greater than 50% made in each condition. By removing these subjects, the average offer removed in each condition was 550, 504, and 565 in the low, middle, and high stakes. Removal of these subjects does not affect the results in any significant way. The subjects removed were subjects number 301, 405, 506, 810, 904, and 1004. (These subjects are included in the summary statistics in Table I (top row) and are in Appendix A.)

[28] The rejection of .5% in all but the eighth period can be seen in Figure 1b in the 0–245 offer range.

where PIE captures the three stakes levels, $ROUND$ represents the (linear) amount of experience a player has ($ROUND = 1$ in round 1, etc.), $SUB(PIE)$ captures the (dependent) fixed subject effects, noting that subjects are nested within a single PIE treatment, and $PIE * ROUND$ captures any unique interaction between experience and stakes effects.[29]

Table IV summarizes the results and Figure 3b shows the predicted offers from the model. There is a significant interaction between stakes and round between the middle and low stakes conditions ($F = 10.30$, $p < .01$) and a marginally significant interaction between stakes and round for the middle and high stakes conditions ($F = 2.94$, $p < .10$). Middle stakes offers are decreasing more than either the low or high stakes conditions (Figure 3b shows this steeper slope). Because of this interaction, we cannot investigate a main effect between the middle stakes and the other two conditions.[30] However, comparing the high and low stakes conditions, where no interaction occurs, we cannot reject that high stakes offers are the same as low stakes offers ($f = 1.14$, $p > .20$).

Although stakes have no main effect on offers, offers decreased significantly more in the middle than in the low stakes. We now explore whether the different learning patterns across treatments can be explained by initial differ-

TABLE IV

ANOVA RESULTS:

PIE SIZE (STAKES) AND INTERACTION OF PIE AND ROUND EFFECTS ON OFFER

Contrasts	PIE	PIE*ROUND
middle vs. low	$F_{1,50} = 2.82$, $p < .10$	$F_{1,50} = 10.30$, $p < .01$
high vs. low	$F_{1,44} = 1.14$, $p > .25$	$F_{1,44} = 2.00$, $p > .15$
high vs. middle	$F_{1,50} = 7.87$, $p < .01$	$F_{1,50} = 2.94$, $p < .10$

Notes: Analysis of Variance Model: $OFFER = PIE + ROUND + SUB(PIE) + PIE^{*}ROUND$. The model predicts, for the average proposer within each treatment, the following:

Pie = 60: $OFFER = 440 + 0.07^{*}ROUND$,
Pie = 300: $OFFER = 453 - 5.16^{*}ROUND$,
Pie = 1500: $OFFER = 423 - 2.13^{*}ROUND$,

or

$$OFFER = 440 + .07^{*}ROUND + 13^{*}PIE_m - 17^{*}PIE_h - 5.23^{*}PIE_m{}^{*}ROUND - 2.13^{*}PIE_h{}^{*}ROUND.$$
$$(p > .25) \qquad (p < .01) \qquad (p > .15)$$

[29] For a detailed description of analysis of variance, see Winer (1971). The ANOVA model assumes experience has a linear effect on offers. Although experience may have nonlinear effects on offers, we found no significant differences between the linear and several nonlinear models. The linear model has out-of-sample concerns, such as suggesting that offers in high rounds (e.g., rounds greater than 1000) may be greater or less than the size of the pie. We limit our conclusions to the scope of the ten rounds of the experiment and do not extrapolate beyond them.

[30] It is meaningless to talk about an overall difference between offers in the middle stakes and the other two conditions because the interaction signifies that the effect of stakes depends critically on the amount of experience. This result can be seen in Figures 3a or 3b where middle stakes offers are falling relative to low stakes offers; in early rounds middle and low stakes offers are similar, but in later rounds middle stakes offers are lower.

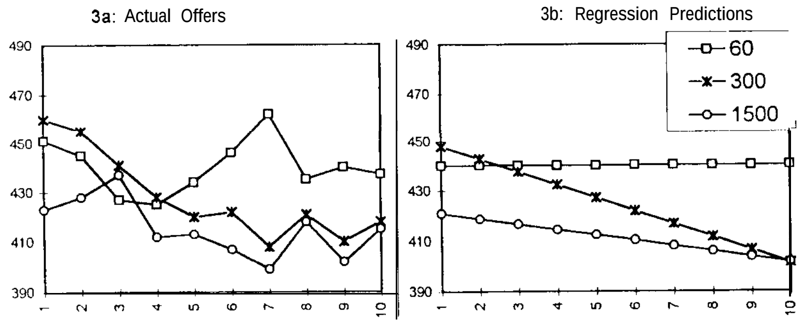


FIGURE 3a–3b.—Subjects who made no more than two offers > 50%.

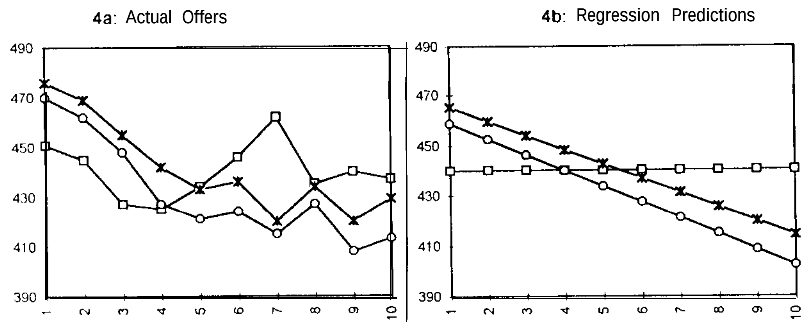
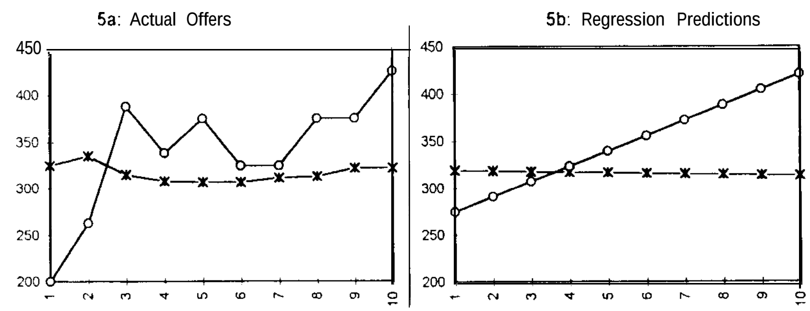


FIGURE 4a–4b.—Subjects whose round 1 offer is at least 350 (and who made no more than two offers > 50%).



Notes: **Figures 3a - 5b:** Squares = 60 Sk, X's = 300 Sk and Circles = 1500 Sk

FIGURE 5a–5b.—Subjects whose round 1 offer is less than 350 (and who made no more than two offers > 50%).

ences across stakes among proposers. One potentially important difference among inexperienced proposers is that no proposer in the low stakes made an offer below 35% of the pie in the first round, whereas seven proposers in the higher two conditions made offers less than 35%. One hypothesis is that these initial differences rather than differences among responders could cause the different learning patterns.

Figures 4a and 5a separate the behavior of proposers who in round 1 made an offer of at least 35% (4a) from those who made an offer less than 35% (5a). Figures 4b and 5b plot regression results (model 7) for these offers. Figure 4b shows that average offers in the higher two stakes conditions fall over time while there is no change in offers in the low stakes condition when round 1 offers are at least 35%. The interaction between round and pie size is highly significant ($F > 15, p < .0001$ for both middle vs. low and high vs. low comparisons) and there is no difference between the two higher stakes conditions ($F = 0.14$, $p > .40$). Thus, when proposers initially made similar offers across stakes (defined here as offers of at least 35% in the first round), higher stakes proposers decreased their offers more than low stakes proposers, indicating that initial differences among proposers cannot explain the different observed learning patterns.

Figures 5a and 5b show that high stakes proposers who initially make relatively small offers increase their offers compared to middle stakes proposers.[31] Comparing Figures 3b, 4b, and 5b, the few proposers who increased their average offers in the highest stakes condition (Figure 5b) explain why the overall average offers in the highest stakes do not decrease much: these few proposers in early rounds bring down and in later rounds bring up the average offer of all high stakes proposers. In the middle stakes condition, however, proposers who initially made low offers (less than 35%) continued to make relatively low offers (less than 35 %) and hence did not retard the overall average offer from falling over time.

4. LEARNING

The current results indicate that offers by inexperienced subjects are alike across stakes, but become different with experience. This is similar to that observed by Roth et al. (1991) in comparing different subject pools. The Roth and Erev (1995) reinforcement learning model was successfully used to predict the different learning behavior observed in those experiments. If the learning model can also predict the different learning behavior in the different stakes conditions in the current experiment, then one question the learning model can address is whether the initial differences in proposer behavior or the differences

[31] Since only 7 subjects made offers less than 35% in period 1 in the higher two stakes conditions, statistical analysis of their offers is omitted.

in responder behavior can explain the different learning patterns across the stakes treatments.

The reinforcement learning model assumes each player has an initial propensity to play each of a finite number of pure strategies (see Roth and Erev for a full description of the model). The propensity to play each pure strategy is updated (reinforced) each time the strategy is played, by adding the monetary payoff just earned to the current propensity to play the strategy. For each subject, the probability of playing a strategy equals the propensity to play the strategy divided by the sum of the propensities of all the strategies. The learning model is investigated by having simulated proposers and responders play each other in a simulation of the experimental environment. For brevity we omit the details of the simulations we have run of the current experiment.

We used the behavior of experimental proposers and responders within the first two rounds of each treatment to generate initial propensities for simulated proposers and responders.[32] With these initial propensities, 5,000 simulations were run for each treatment. Although simulated offers changed more slowly than experimental offers, the direction of learning for each treatment was the same for simulated and experimental offers. Consistent with the experimental results. simulated middle stakes offers decreased most, highest stakes offers decreased second most, and lowest stakes offers decreased least.

We next explored whether the different learning patterns across treatments can be explained by initial differences across stakes among proposers or by the lower likelihood of rejection in higher stakes among responders. The simulation results show that no matter what the initial propensities of proposers, the change in offers over time depends critically on the responders they played against. If proposers play against lower stakes responders, offers fall the least (increase the most) relative to playing against either middle or high stakes responders. The learning model thus suggests that the different learning behavior observed is the result of the lower rejection rates observed in the higher stakes; all simulated proposers learn to lower offers when playing against middle and high stakes responders while they all learn to increase offers when playing against low stakes responders.[33]

5. CONCLUSIONS

Our experimental results for hoth the market and ultimatum games support the conclusion that, both when observed behavior conforms to perfect equilibrium predictions and when it does not, behavior of inexperienced players may be robust to large increases in rewards. Our ultimatum game results confirm prior experimental results in this regard, while in other respects they considerably extend what has previously been observed.

[32] Roth and Erev (1995) describe the process used to determine initial propensities. See also Erev and Roth (1998).

[33] The learning model results are reported in Slonim and Roth (1996).

As discussed earlier, a number of experiments have now established the fact that single-play ultimatum game behavior is quite robust, and does not approach the perfect equilibrium predictions (for either player) even when stakes are quite high. Perhaps the most compelling of these is the experiment of Cameron (1995), which detected no change in behavior even in the face of a change in stakes by a factor of 40. Our results are quite consistent with this: in round 1, behavior in all three of our treatments is quite similar, and far from the perfect equilibrium predictions.

Of course the failure to detect statistically significant differences does not mean that not even small differences exist. Variables like rejection frequency present a particularly difficult case, since only the smaller observed offers are rejected with high frequency, and such offers are rare, so that trying to detect differences in first-round rejection rates would require impractically large samples. The learning model of Roth and Erev (1995) predicts that small initial differences in rejection frequencies should be reflected in increasingly different proposals as players have an opportunity to learn about the game, and the experiment reported here was designed to explore this prediction.

Two differences in the ultimatum game behavior were detected as stakes increased. First, responders (pooled over all rounds) rejected offers less often. Second, there was an interaction effect between stakes and experience: in the higher stakes conditions the offers decreased with experience. The experiment and learning simulations suggest that small initial differences in proposer behavior cannot account for the differential learning behavior, but that the lower likelihood of being rejected in the higher stakes can account for higher stakes proposers learning to make lower offers.

Notice that the different patterns of learning we observe among proposers in the different stakes conditions of the experiment, and the hypothesis about its origin in the different rejection frequencies which the learning model provides, tell us something about rejection frequencies which the simple statistical analysis cannot. Not only are the differences in rejection frequencies across stakes statistically significant, apparently they are also behaviorally important.

In general, new kinds of theory allow us to explore different kinds of questions, and suggest different kinds of experiments. We therefore view this paper not only as an experiment designed to explore the effects of large changes in stakes, but also as an attempt to take seriously the demands that theories of learning place on (and the opportunities they provide for) experimental design and analysis.

Dept. of Economics, University of Pittsburgh, Pittsburgh, PA 15260, U.S.A.; slonim + @pitt.edu

and

Dept. of Economics, University of Pittsburgh, Pittsburgh, PA 15260, U.S.A.; alroth + @pitt.edu; http: // www.pitt.edu / ~alroth.html

Manuscript received October, 1996; final revision received June, 1997.

APPENDIX:
INDIVIDUAL ULTIMATUM GAME BEHAVIOR

Low Pie Stakes (60 Slovak Crowns)[a]

Pro#	PIE	oP	r	rs#	o1	r	rs#	o2	r	rs#	o3	r	rs#	o4	r	rs#	o5	r	rs#	o6	r	rs#	o7	r	rs#	o8	r	rs#	o9	r	rs#	o10	r	rs#
201	60	420	0	216	400	1	219	470	0	213	400	0	220	480	1	211	500	0	216	470	0	212	500	0	218	420	0	215	500	0	220	550	0	211
202	60	460	1	215	500	0	211	500	0	216	500	0	212	500	0	218	500	0	215	500	0	219	500	0	213	500	0	220	500	0	212	550	0	218
203	60	450	0	220	430	0	218	420	0	215	415	0	219	410	1	213	420	0	220	420	0	211	420	0	216	420	0	212	415	1	219	420	0	213
204	60	500	0	218	500	0	220	490	0	211	500	0	216	485	0	212	490	0	218	500	0	215	500	0	219	485	0	213	485	0	216	485	0	212
205	60	495	0	219	400	0	216	365	0	212	250	0	218	250	1	215	335	1	219	400	1	213	335	0	220	335	0	211	335	0	218	335	0	215
206	60	480	0	211	475	1	215	480	0	219	480	0	213	475	0	220	470	1	211	475	0	216	470	0	212	475	0	218	475	0	213	470	0	220
208	60	500	0	212	500	0	213	500	0	220	500	0	211	500	0	216	500	0	212	500	0	218	500	0	215	500	0	219	500	0	211	500	0	216
210	60	500	0	213	500	0	212	500	0	218	500	0	215	500	0	219	500	0	213	500	0	220	500	0	211	500	0	216	500	0	215	500	0	219
301	60	600	0	311	515	0	314	550	1	319	600	0	313	575	0	320	550	0	311	550	0	316	550	0	312	550	0	318	530	0	315	525	0	313
302	60	500	0	315	500	0	311	400	1	316	450	0	312	500	0	318	400	0	315	350	0	314	600	0	319	300	1	313	350	0	320	350	0	312
303	60	450	0	313	470	0	318	470	0	315	430	1	314	470	1	319	485	0	313	485	0	320	485	0	311	470	0	316	470	0	312	470	0	314
304	60	500	0	312	500	0	313	500	0	320	500	0	311	500	0	316	500	0	312	500	0	318	500	0	315	500	0	314	500	0	319	400	1	311
305	60	415	0	314	400	0	316	380	0	312	380	0	318	330	0	315	250	1	314	380	1	319	500	0	313	450	0	320	400	1	311	450	0	318
306	60	500	0	320	465	0	315	395	0	314	365	1	319	370	1	313	390	0	320	380	1	311	390	1	316	395	0	312	395	0	318	395	1	319
307	60	450	0	318	420	0	320	410	0	311	400	1	316	400	0	312	400	0	318	385	0	315	385	1	314	400	1	319	400	1	313	410	0	316
308	60	500	0	316	500	0	319	450	1	313	500	0	320	495	0	311	485	0	316	480	0	312	490	0	318	475	0	315	470	0	314	470	0	320
310	60	500	0	319	500	0	312	420	0	318	350	0	315	350	1	314	335	1	319	420	1	313	420	0	320	420	0	311	420	0	316	420	0	315
701	60	400	0	712	430	0	713	400	1	720	450	0	711	475	0	716	455	0	712	470	0	718	460	0	715	465	0	720	450	0	711	445	0	716
703	60	400	0	711	400	0	718	425	0	715	400	0	713	455	0	720	455	0	711	400	0	716	400	0	712	400	0	715	400	0	713	400	0	720
704	60	500	0	715	400	0	711	300	1	716	340	0	712	350	0	718	350	0	715	650	0	713	400	0	720	350	1	716	350	1	712	400	0	718
705	60	500	0	713	400	0	716	500	0	712	200	1	718	300	1	715	500	0	713	400	0	720	375	0	711	400	0	712	400	1	718	400	0	715
706	60	550	0	716	435	0	715	455	0	713	500	0	720	425	0	711	520	0	716	450	0	712	500	0	718	450	0	713	510	0	720	390	0	711
708	60	650	0	718	350	1	720	500	0	711	500	0	716	500	0	712	500	0	718	500	0	715	500	0	713	400	0	711	400	0	716	400	0	712
710	60	750	0	720	500	0	712	500	0	718	500	0	715	250	0	713	250	1	720	250	1	711	500	0	716	500	0	718	500	0	715	500	0	713

[a]Notation: Pro# = Proposer Number, PIE = Slovak Crowns Size of Pie, oP = offer in Practice Round, o1 ... o10 = offers in rounds 1 ... 10, r = 0: offer was accepted, r = 1: offer was rejected, rs# = responder number.

Middle Pie Stakes (300 Slovak Crowns)[a]

Pro#	PIE	oP	r	rs#	o1	r	rs#	o2	r	rs#	o3	r	rs#	o4	r	rs#	o5	r	rs#	o6	r	rs#	o7	r	rs#	o8	r	rs#	o9	r	rs#	o10	r	rs#
401	300	5	1	418	5	1	413	5	1	420	5	1	411	5	1	416	5	1	412	5	1	418	5	1	415	5	0	419	5	1	420	5	1	411
402	300	400	1	413	490	0	416	500	0	412	495	0	418	505	0	415	490	0	419	500	0	413	490	0	420	480	0	411	480	0	412	500	0	418
403	300	500	0	416	415	0	415	430	0	419	400	0	413	420	0	420	435	0	411	410	0	416	400	0	418	420	0	415	430	0	419	410	0	413
404	300	100	0	419	700	0	411	600	0	416	500	0	412	450	0	418	400	0	415	400	0	419	100	1	413	500	0	420	400	0	416	450	0	412
405	300	500	0	420	500	0	412	450	0	418	300	1	415	670	0	419	550	0	413	535	0	420	500	0	411	520	0	416	500	0	418	480	0	415
406	300	500	0	412	500	0	419	500	0	413	500	0	420	500	0	411	500	0	416	500	0	412	500	0	418	500	0	415	500	0	413	500	0	420
408	300	400	0	415	300	0	420	335	1	411	335	0	416	335	0	412	335	0	418	335	1	415	335	0	419	335	0	413	335	1	411	335	0	416
410	300	500	0	411	450	0	418	450	0	415	440	0	419	440	0	413	430	0	420	420	0	411	400	0	416	390	0	412	350	0	415	335	0	419
501	300	535	0	516	450	0	519	550	0	513	535	0	520	500	0	511	450	0	516	470	0	512	485	0	518	450	1	515	500	0	520	485	0	511
502	300	500	0	515	500	0	511	500	0	516	500	0	512	500	0	518	500	0	515	500	0	519	500	0	513	500	1	520	500	0	512	500	0	518
503	300	400	1	520	420	0	518	415	1	515	420	0	519	420	0	513	395	1	520	420	0	511	420	1	516	420	0	512	420	0	519	420	0	513
504	300	600	0	518	500	0	520	450	0	511	380	0	516	370	1	512	420	0	518	430	1	515	450	0	519	440	0	513	420	0	516	430	0	512
505	300	480	1	519	480	0	516	490	0	512	460	0	518	450	1	515	470	0	519	480	0	513	470	1	520	490	0	511	490	0	518	500	0	515
506	300	500	0	511	550	0	515	400	0	519	550	0	513	500	0	520	400	0	511	400	1	516	660	1	512	450	0	518	450	0	513	700	0	520
508	300	500	0	512	500	0	513	500	0	520	500	0	511	500	0	516	500	0	512	500	0	518	500	0	515	500	0	519	500	0	520	500	0	511
510	300	500	0	513	500	0	512	470	0	518	450	0	515	430	0	519	415	0	513	400	1	520	430	0	511	430	0	516	430	1	518	500	0	515
801	300	460	0	812	460	0	813	450	0	820	360	0	811	360	0	816	350	0	812	340	0	818	340	0	815	270	0	820	270	0	811	270	0	816
803	300	495	0	811	500	0	818	490	0	815	485	0	813	475	0	820	470	0	811	465	0	816	460	0	812	450	0	815	435	0	813	400	0	820
804	300	500	0	815	335	0	811	340	0	816	300	0	812	290	0	818	290	1	815	290	1	813	300	0	820	300	0	816	300	0	812	300	0	818
805	300	30	1	813	370	0	816	390	0	812	435	0	818	340	0	815	340	1	813	395	0	820	395	0	811	390	0	812	380	0	818	380	0	815
806	300	470	0	816	400	0	815	350	0	813	360	0	820	300	0	811	310	0	816	280	0	812	270	0	818	260	1	813	275	0	820	290	0	811
808	300	470	0	818	480	0	820	480	0	811	490	0	816	480	0	812	480	0	818	480	0	815	470	0	813	480	0	811	480	0	816	470	0	812
810	300	500	0	820	470	0	812	505	0	818	490	1	815	520	0	813	500	0	820	515	0	811	510	0	816	525	0	818	505	0	815	515	0	813
1101	300	330	1	1120	490	0	1117	490	0	1114	490	0	1119	480	0	1113	470	0	1120	460	0	1111	460	0	1116	440	0	1112	435	0	1118	430	0	1115
1102	300	500	0	1115	500	0	1111	500	0	1116	450	0	1112	400	0	1118	400	1	1115	400	0	1117	400	0	1114	400	0	1119	350	0	1113	350	0	1120
1103	300	500	0	1119	490	0	1118	480	0	1115	500	0	1117	470	0	1114	500	0	1119	470	0	1113	460	0	1120	500	0	1111	460	0	1116	500	0	1112
1104	300	295	1	1116	340	0	1119	330	0	1113	310	0	1120	300	0	1111	295	1	1116	295	0	1112	300	1	1118	305	1	1115	330	0	1117	330	0	1114
1105	300	500	0	1117	500	0	1116	500	0	1112	500	0	1118	500	0	1115	500	0	1117	500	0	1114	500	0	1119	500	0	1113	500	0	1120	500	0	1111
1106	300	500	0	1113	465	0	1115	460	0	1117	430	0	1114	400	0	1119	370	0	1113	400	0	1120	370	0	1111	370	0	1116	370	0	1112	400	0	1118
1107	300	400	0	1118	350	0	1120	350	0	1111	350	1	1116	400	0	1112	400	0	1118	400	1	1115	360	0	1117	360	0	1114	360	0	1119	360	0	1113
1108	300	500	0	1111	465	0	1114	400	0	1119	400	0	1113	400	0	1120	330	0	1111	330	0	1116	400	0	1112	400	0	1118	400	1	1115	400	0	1117
1109	300	500	0	1112	500	0	1113	500	0	1120	500	0	1111	500	0	1116	470	0	1112	480	0	1118	440	0	1115	440	0	1117	440	0	1114	470	0	1119
1110	300	500	0	1114	500	0	1112	500	0	1118	505	0	1115	505	0	1117	470	0	1114	500	0	1119	440	0	1113	500	0	1120	350	0	1111	405	0	1116

[a]Notation: Pro# = Proposer Number, PIE = Slovak Crowns Size of Pie, oP = offer in Practice Round, o1 ... o10 = offers in rounds 1 ... 10, r = 0: offer was accepted, r = 1: offer was rejected, rs# = responder number.

High Pie Stakes (1500 *Slovak Crowns*)[a]

Pro#	PIE	oP	r	rs#	o1	r	rs#	o2	r	rs#	o3	r	rs#	o4	r	rs#	o5	r	rs#	o6	r	rs#	o7	r	rs#	o8	r	rs#	o9	r	rs#	o10	r	rs#
601	1500	500	0	611	750	0	614	400	0	619	400	0	613	330	0	620	250	0	611	250	0	616	200	1	612	500	0	618	330	0	615	330	0	620
602	1500	750	0	615	250	0	611	100	0	616	600	0	612	150	1	618	400	0	615	300	0	614	200	1	619	400	0	613	400	0	620	400	0	618
603	1500	400	1	613	400	1	618	500	0	615	500	0	614	500	0	619	500	0	613	480	0	620	480	0	611	470	0	616	400	0	612	400	0	619
604	1500	999	0	612	100	1	613	500	0	620	500	0	611	500	0	616	400	0	612	300	1	618	400	0	615	400	0	614	400	0	619	400	0	616
605	1500	350	1	614	400	0	616	400	0	612	400	0	618	400	0	615	400	0	614	400	1	619	400	0	613	400	1	620	400	0	611	400	0	615
606	1500	400	0	620	450	0	615	460	0	614	400	0	619	420	0	613	410	0	620	400	0	611	410	0	616	400	0	612	380	0	618	400	0	613
607	1500	450	0	618	450	0	620	480	0	611	440	0	616	400	0	612	430	0	618	425	0	615	435	0	614	440	0	619	440	0	613	440	0	612
608	1500	500	0	616	250	1	619	250	0	613	250	1	620	500	0	611	500	0	616	500	0	612	500	0	618	500	0	615	500	0	614	500	0	611
610	1500	400	1	619	470	0	612	465	0	618	460	0	615	450	0	614	440	1	619	440	0	613	440	0	620	430	0	611	440	0	616	430	0	614
901	1500	500	0	912	500	0	913	580	0	920	500	0	911	580	0	916	500	0	912	500	0	918	500	0	915	500	0	920	500	0	911	500	0	916
903	1500	500	0	911	500	0	918	500	0	915	500	0	913	500	0	920	500	0	911	500	0	916	500	0	912	500	0	915	500	0	913	500	0	920
904	1500	435	0	915	800	0	911	670	0	916	600	0	912	730	0	918	500	0	915	730	0	913	730	0	920	800	0	916	500	0	912	665	0	918
905	1500	600	0	913	500	0	916	470	0	912	450	0	918	400	0	915	400	0	913	400	0	920	370	0	911	370	0	912	370	0	918	335	0	915
906	1500	400	0	916	410	0	915	330	0	913	380	0	920	200	0	911	270	1	916	400	0	912	340	0	918	340	0	913	340	0	920	330	0	911
908	1500	500	0	918	500	0	920	600	0	911	600	0	916	500	0	912	500	0	918	500	0	915	500	0	913	500	0	911	500	0	916	500	0	912
910	1500	500	0	920	500	0	912	500	0	918	500	0	915	500	0	913	500	0	920	500	0	911	500	0	916	500	0	918	500	0	915	500	0	913
1001	1500	480	0	1011	410	0	1014	400	0	1019	380	0	1013	370	0	1020	370	0	1011	360	0	1016	350	0	1012	340	1	1018	340	0	1015	350	0	1013
1002	1500	400	1	1015	400	0	1011	400	0	1016	400	0	1012	400	0	1018	400	0	1015	400	0	1014	400	0	1019	400	0	1013	400	0	1020	400	0	1012
1003	1500	200	0	1013	200	1	1018	200	1	1015	200	0	1014	200	1	1019	200	0	1013	200	0	1020	200	1	1011	200	0	1016	200	0	1012	400	0	1014
1004	1500	400	1	1012	500	0	1013	610	0	1020	345	0	1011	710	0	1016	240	1	1012	220	1	1018	490	0	1015	505	0	1014	360	0	1019	600	0	1011
1005	1500	500	0	1014	500	0	1016	500	0	1012	500	0	1018	500	0	1015	500	0	1014	500	0	1019	500	0	1013	500	0	1020	500	0	1011	500	1	1018
1006	1500	470	0	1020	435	0	1015	500	0	1014	400	0	1019	370	0	1013	335	0	1020	300	0	1011	260	0	1016	220	0	1012	170	1	1018	200	1	1019
1007	1500	500	0	1018	460	0	1020	400	0	1011	400	0	1016	400	0	1012	400	0	1018	400	0	1015	400	0	1014	400	0	1019	340	0	1013	340	0	1016
1008	1500	400	0	1016	400	0	1019	400	0	1013	400	0	1020	400	0	1011	400	0	1016	400	0	1012	400	0	1018	400	0	1015	400	0	1014	500	0	1020
1010	1500	500	0	1019	500	0	1012	500	0	1018	500	0	1015	500	0	1014	500	0	1019	500	0	1013	500	0	1020	500	1	1011	500	0	1016	500	0	1015

[a] Notation: Pro# = Proposer Number, PIE = Slovak Crowns Size of Pie, oP = offer in Practice Round, o1 … o10 = offers in rounds 1 … 10, r = 0: offer was accepted, r = 1: offer was rejected, rs# = responder number.

REFERENCES

BERK, JONATHON, ERIC HUGHSON, AND KIRK VANDEZANDE (1996): "The Price is Right, But Are the Bids? An Investigation of Rational Decision Theory," *American Economic Review*, 86, 954–970.

BINSWANGER, HANS P. (1980): "Attitudes toward Risk: Experimental Measurement In Rural India," *American Journal of Agricultural Economics*, 62, 395–407.

BOLTON, GARY (1991): "A Comparative Model of Bargaining: Theory and Evidence," *American Economic Review*, 81, 1096–1136.

BOLTON, GARY E., AND RAMI ZWICK (1995): "Anonymity versus Punishment in Ultimatum Bargaining," *Games and Economic Behavior*, 10, 95–121.

CAMERON, LISA (1995): "Raising the Stakes in the Ultimatum Game: Experimental Evidence from Indonesia," Working Paper #345, Industrial Relations Section, Princeton University.

EREV, IDO, AND ALVIN E. ROTH (1998): "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique Mixed Strategy Equilibria," *American Economic Review*, forthcoming.

FEHR, ERNST, AND ELENA TOUGAREVA (1995): "Do Competitive Markets with High Stakes Remove Reciprocal Fairness: Experimental Evidence from Russia," mimeo, Institute for Empirical Research in Economics, University of Zurich.

FUDENBERG, DREW, AND DAVID LEVINE (1997): "Measuring Players' Losses in Experimental Games," *Quarterly Journal of Economics*, 112, 507–536.

GERTNER, ROBERT (1993): "Game Shows and Economic Behavior: Risk-Taking on 'Card Sharks'," *Quarterly Journal of Economics*, 108, 507–522.

GUTH, WERNER, R. SCHMITTBERGER, AND R. SCHWARTZ (1982): "An Experimental Analysis of Ultimatum Bargaining," *Journal of Economic Behavior and Organization*, 3, 367–388.

GUTH, WERNER, AND REINHARD TIETZ (1990): "Ultimatum Bargaining Behavior: A Survey and Comparison of Experimental Results," *Journal of Economic Psychology*, 11, 417–449.

HOFFMAN, E., K. A. MCCABE, AND V. L. SMITH (1996): "On Expectations and the Monetary Stakes in Ultimatum Games," *International Journal of Game Theory*, 25, 289–302.

KACHELMEIER, STEVEN J., AND MOHAMED SHEHATA (1992): "Examining Risk Preferences Under High Monetary Incentives: Experimental Evidence from the People's Republic of China," *American Economic Review*, 82, 1120–1141.

METRIC, ANDREW (1995): "A Natural Experiment in Jeopardy," *American Economic Review*, 85, 240–253.

OCHS, JACK, AND ALVIN E. ROTH (1989): "An Experimental Study of Sequential Bargaining," *American Economic Review*, 79, 355–384.

ROTH, ALVIN E. (1995a): "Introduction to Experimental Economics," *Handbook of Experimental Economics*, ed. by John Kagel and Alvin E. Roth. Princeton: Princeton University Press, 3–109.

——— (1995b): "Bargaining Experiments," *Handbook of Experimental Economics*, ed. by John Kagel and Alvin E. Roth. Princeton: Princeton University Press, 253–348.

ROTH, ALVIN E., AND IDO EREV (1995): "Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term," *Games and Economic Behavior*, 8, 164–212.

ROTH, ALVIN E., VESNA PRASNIKAR, MASAHIRO OKUNO-FUJIWARA, AND SHMUEL ZAMIR (1991): "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study," *American Economic Review*, 81, 1068–1095.

SLONIM, ROBERT L. (1995): "Learning, Bounded Rationality and Financial Incentives: An Experimental Investigation," 1995 Dissertation, Duke University.

——— (1997): "Rejecting Rejections: A Comment on the Power of Testing Ultimatum Game Responder Behavior," Mimeo, University of Pittsburgh.

SLONIM, ROBERT L., AND ALVIN E. ROTH (1996): "Learning and Financial Incentives in a Market and Ultimatum Game: An Experiment in the Slovak Republic," Mimeo, University of Pittsburgh.

STRAUB, PAUL G., AND J. KEITH MURNIGHAN (1995): "An Experimental Investigation of Ultimatum Games: Common Knowledge, Fairness, Expectations, and Lowest Acceptable Offers," *Journal of Economic Behavior and Organization*, 27, 345–364.

TELSER, L. G. (1995): "The Ultimatum Game and Law of Demand," *The Economic Journal*, 105, 1519–1523.

WALLIS, W. ALLEN, AND MILTON FRIEDMAN (1942): "The Empirical Derivation of Indifference Functions," in *Studies in Mathematical Economics and Econometrics in Memory of Henry Schultz*, ed. by O. Lange, F. McIntyre, and T. O. Yntema. Chicago: University of Chicago Press, 175–189.

WINER, B. J. (1971): *Statistical Principles of Experimental Design*, Second Edition. New York: McGraw-Hill.

Econometrica, Vol. 66, No. 3 (May, 1998), 597–626

COMMUNICATION IN REPEATED GAMES WITH IMPERFECT PRIVATE MONITORING

BY OLIVIER COMPTE[1]

This paper examines repeated games in which each player observes a private and imperfect signal on the actions played, and in which players are allowed to communicate using public messages. Providing incentives for players to *reveal* their observations generates (revelation) constraints that, combined with signal imperfections, may be a source of *inefficiencies*. However, by *delaying* the revelation of their observations, players may economize on the cost of deterring deviations, and thereby avoid these inefficiencies.

Because a player would not want to trigger a sanction that would penalize him too, revelation constraints also tend to make sanctions difficult to enforce. However, with at least three players, detecting deviations may not require that *all* the players reveal their observations. In that case, we obtain a Nash threat version of the Folk theorem. With two players, we do not obtain a similar result. Nevertheless, we show that an efficient outcome can (almost) always be approximated.

KEYWORDS: Repeated games, imperfect private monitoring, delayed communication.

1. INTRODUCTION

THIS PAPER EXAMINES REPEATED GAMES in which each player observes a private and imperfect signal on the actions played. Compte[2] (1994) and Kandori and Matsushima (1994) have shown that in this class of games, allowing players to communicate using public messages is useful because it allows players to coordinate their behavior. The focus of the present paper is different. Private signals have the feature that players may choose *when* to make them public, and our purpose is to analyze if and when *delaying communication* helps players to support efficient outcomes.

A well-known application of repeated games is the analysis of collusion in repeated oligopoly (Green and Porter (1984), Abreu, Pearce, and Stacchetti (1986)). In these papers, as well as in many other studies, players' observations are assumed to be public.[3] However, in some situations of interest, players only receive private signals. In Stigler's (1964) secret price cutting model, for exam-

[1] I am especially grateful to Paul Milgrom for his guidance. I thank Faruk Gul, Peter Hammond, Philippe Jéhiel, the editor, and two anonymous referees for detailed comments, and Ennio Stacchetti for his early encouragement. I also thank seminar participants at Stanford, CEPREMAP, and "séminaire GREMAQ" at University of Toulouse. This paper is based on Chapter 3 of my Ph.D. dissertation at Stanford.

[2] Compte (1994) is an earlier version of this paper.

[3] These studies include the classical papers by Aumann and Shapley (1976) and Rubinstein (1979), as well as more recent ones by Abreu, Pearce, and Stacchetti (1990) and Fudenberg, Levine, and Maskin (1994).