Predicting the Role of Fairness in Bargaining

Out –of-Sample Estimates from a Social Utiling
Model with Quantal Response

by
Arnaud De Bruyn
Pennsylvania State University
Gary E. Bolton
Pennsylvania State University

ISBM Report 1-2004

Institute for the Study of Business Markets
The Pennsylvania State University
402 Business Administration Building
University Park, PA 16802-3004
(814) 863-2782 or (814) 863-0413 Fax

Predicting the Role of Fairness in Bargaining

Out-of-Sample Estimates from a Social Utility Model with Quantal Response

Arnaud De Bruyn & Gary E. Bolton[♣]

Penn State University

February 2004

We fit a model to 1-round sequential (ultimatum) bargaining game data, and use the model to obtain out-of-sample estimates of behavior in multiple round sequential bargaining games. The model embeds a social utility function in a quantal response framework, and has 3 fitted parameters, 1 to capture utility trade-offs and 2 to capture experience effects. The data used for out-of-sample testing comes from 6 previously reported studies, encompassing 20 distinct parameterizations of the sequential bargaining game. Out-of-sample estimates account for about 60% of the variance of average first offers, 54% of rejection behavior (save for games where the pie size is very small), and 67% of disadvantageous counteroffers. By and large, with experience, bargaining behavior moves towards the static social utility model equilibrium. Alternative fits of the model indicate that parameter estimates are stable.

[♣] De Bruyn: 701 Business Administration Building, Smeal College of Business, Penn State University, University Park, PA 16802. Tel.: (+1) (814) 865-4091. Fax: (+1) (814) 865-3015, adebruyn@psu.edu. Bolton: 310 Business Administration Building, Smeal College of Business, Penn State University, University Park, PA 16802, Tel.: (+1) (814) 865-0611, gbolton@psu.edu.

1 INTRODUCTION

This paper investigates the ability of a social utility model to predict bargaining behavior out-of-sample. The model admits a preference for fairness, the influence of which on bargaining has been extensively documented in laboratory studies (Roth, 1995), and is commonly thought to be an influence on bargaining in the field.[1] Social preference models are an emerging body of theory that captures many of the ordinal regularities associated with fairness. But whether these models are sufficiently robust for the more exacting task of quantitative prediction remains an open question. One important reason to wonder is that people are known to differ in their assessments of what-is-fair; and such assessments are arguably situation specific. Given this, it is unclear whether a portable model can be expected to yield reliable forecasts.

We focus on sequential bargaining, a class of games that has for some time been the dominant paradigm for the study of bargaining in economics. We fit a social utility model cast in a quantal response framework (thus permitting "noisy" choices) to data taken from a well known ultimatum game study. We then derive fully out-of-sample estimates of behavior for a series of multiple round sequential bargaining experiments, performed by various investigators. The ultimatum game is a single round sequential bargaining game. In essence, we fit the model from the simplest member of the class to see how well the model predicts across the rest of the class. The results indicate that a model that makes relatively simple adjustments to the conventional subgame perfect equilibrium theory can provide quite accurate out-of-sample estimates of bargaining behavior; and in situations where conventional theory does quite poorly.

Questions about the robustness of fairness were one of the initial sparks to empirical research into multiple round sequential bargaining. Speaking of Güth et al.'s (1982) seminal study of the ultimatum game, Binmore, Shaked and Sutton (1985) framed the issue this way

> *Our suspicion is that the one-stage ultimatum game is a rather special case, from which it is dangerous to draw general conclusions. In the ultimatum game, the first player might be dissuaded from making an opening offer at, or close to, the "optimum" level, because his opponent would then incur a negligible cost from making an "irrational" rejection. In the two-stage game, these considerations are postponed to the second-stage, and so their impact is attenuated. p.1180*

Binmore et al.'s subsequent experiment on a two-round version of the game ignited the series of studies that we revisit here. We examine all the baseline studies of multiple round sequential

[1] To give a thumbnail statistic, a recent Google search on 'fair bargaining' yielded some 441,000 links, with business, union, government, press, religious and academic sites all substantially represented.

bargaining overviewed in the "Bargaining" chapter of the *Handbook of Experimental Economics* (Roth, 1995, pp.256-269).

Predicting this data out-of-sample challenges the robustness of the model along three important dimensions. First, the differences in data across studies is such that the researchers involved reached markedly different conclusions concerning the ability of game theory – and fairness – to predict bargaining behavior. The studies differ substantially with respect to procedures and subject pools, features that plausibly influence judgments about what-is-fair – if such judgments are fragile. The simple model we study ignores these considerations, attributing differences, instead, to differences in game parameterization (ex., discount factors, number of rounds, etc.). Second, the data includes observations from 3 and 5 round games, where people have been shown to have difficulty doing the backward induction assumed by the model (Johnson, Camerer, and Sen, 2002, Binmore, McCarthy, Ponti, Samuelson and Shaked, 2002). Our model ignores limitations to backward induction; the sample permits us to assess how much forecast accuracy is lost in doing so. Third, the sample includes experiments in which bargainer behavior changes with bargaining experience. Analysts have typically supposed that static social utility models describe the behavior of experienced players. Our model permits us to assess whether experience, in fact, drives behavior towards the static model equilibrium; if, in fact, experience alters preferences for fairness, we would not expect such to be the case.

One of the major obstacles to quantitatively fitting social utility models is dealing with the experience effects that are common in studies where bargaining is repeated; it is difficult to get accurate estimates of the influence of fairness without separating, and so estimating, the influence of experience. Beginning with Bolton (1991), social utility models have explained sequential bargaining behavior in terms of fairness but typically have not dealt with experience. Consequently, with a few exceptions (discussed below), the predictions stemming from these models have been ordinal in nature. Beginning with Roth and Erev (1995), learning models of ultimatum bargaining have quantified the experience effects. But these models have typically not attempted to capture the role of fairness.

We capture fairness using an ERC (*E*quity-*R*eciprocity-*C*ompetition) social utility specification (Bolton and Ockenfels, 2000). Bargainers have preferences for absolute (pecuniary) and relative (fairness) payoffs, with relative payoffs characterized by a loss of utility as bargaining outcomes move away from equity. The specification has the property that the marginal utility of relative payoff increases as the allocation moves away from an equal split, so that, at the margin, fairness is a bigger issue the further away the allocation is from an equal split. The ERC model is incomplete information with respect to utility payoffs, an important feature of the actual bargaining games. We suspect that other social utility models (e.g., Fehr and Schmidt, 1999) would yield similar results, but do not pursue the issue further here; an in-depth comparison with the many social utility formulations now in the literature requiring, in itself, a lengthily investigation. The model attributes differing assessments of what-is-fair to differing willingness to trade-off absolute and relative payoffs. We estimate a simple version of the model that estimates the average trade-off. Along the same lines, the model focuses on preferences for distribution-based fairness and does not attempt to deal with intention-based fairness effects (Bereby-Meyer and Niederle, forthcoming, Bolton and Ockenfels, forthcoming).

We capture the noise in behavior, as well as experience effects, using McKelvey and Palfrey's (1995) quantal response equilibrium framework. One of the basic assumptions of most choice theories proposed in psychology is that choice behavior is probabilistic (see for instance Luce 1959). Consistent with this, quantal response permits "mistakes" with respect to the optimal decision. Quantal response models has been used to fit decision behavior in a number of studies (ex., Goeree, Holt and Laury, 2000). Haile, Hortaçsu and Kosenok (2003) critique the way quantal response has typically been tested in the literature, and argue that out-of-sample testing is a more appropriate method. They also note that there is little evidence that noise perturbations are stable with subject experience. We model the influence of experience on noise perturbations. The working hypothesis behind our model, which can be checked out-of-sample, is that noise diminishes with experience, moving behavior closer to the static ERC model equilibrium. We emphasize that we do not consider the quantal response framework we employ to be a model of learning, rather we think of it as a relatively straightforward technique for accounting for noise and experience, to permit an estimate of the social utility model.

One of the important issues surrounding bargaining game experiments has been whether rejection behavior might subside for larger bargaining pies. For the range of pie sizes we consider, many studies find no effect, and others but a modest effect.[2] Still, the suspicion

---

[2] Slonim and Roth (1998) who, in an ultimatum game, find a modest sized effect, and then only with substantial player experience, note that no study prior to theirs reliably finds an effect; see their paper for references. Also see Cooper, Feltovich,

persists that stake size influences behavior. Our model supposes that bargainer "certainty of choice" is larger with larger stakes. That is, we suppose there is more noise in a decision when the stakes are small. We will see that this characterization of choice fits the data fairly well save for small stakes where rejection behavior is not as arbitrary as the model suggests.

We fit the model using data from the well known Roth, Prasnikar, Okuno-Fujiwara and Zamir (1991) ultimatum game study. This data set is unique in its combination of size and breadth, characteristics which, as we will explain, are essential to fitting our model. This data has been used in other studies (ex., Costa-Gomez and Zauner, 2001). In this regard, we emphasize that our principle findings are not dependent on this data set: As we will see, fitting the model with data from a multiple round experiment yields very similar estimates.

As part of the analysis of their sequential bargaining data, Goeree and Holt (2003) fit a quantal response model with an embedded social utility function; Costa-Gomez and Zauner (2001), Fehr and Schmidt (1999), Bolton and Ockenfels (2000) and Charness and Rabin (2002) all include calibration exercises for social utility models. Our study differs from these in that we fit an ultimatum game with the aim of extending the model to out-of-sample estimates for other bargaining games. Andreoni and Miller (2002) estimate utility functions for individuals playing the dictator game. They show that there is a good deal of heterogeneity with respect to individual concerns for fairness. We estimate a single function to see whether this simple formulation is sufficient to obtain reasonable estimates of population-level behavior. We note that, while we have observations for many subjects over multiple studies, there is still not enough within-subject data to fit individual utility functions with any confidence – a fact that is suggestive of the critical role population level models are likely to play in forecasting fairness.

In the next section, we describe the studies and associated data in our sample. In section 3, we describe the model and report the calibration to the ultimatum game data. Section 4 reports out-of-sample estimates and comparisons with the data. Section 5 looks at the robustness of our results in three respects: with respect to alternative fits, with respect to quantal mistakes and the symmetry of the relative payoff loss function, and with respect to a three person bargaining game. Section 6 summarizes our findings.

Roth and Zwick (2003) and Abbink, Bolton, Sadrieh and Tang (2001) for alternative views on whether ultimatum game responders learn to change their behavior with experience.

2 THE BARGAINING DATA

One of the important tasks for a new model is to make sense of data that was heretofore considered problematic (as we will see in this section, the conventional subgame perfect equilibrium model exhibits little quantitative fit). The data covers a wide variety of game parameterizations and lab procedures. All told, there are 2,726 observations from 1,037 participants, collected under 21 different experimental conditions by 7 distinct research teams, spread over 7 countries and 3 continents (figures include a three-person ultimatum game study that we will use later for an additional robustness check). Another important advantage of this data is that it was not gathered by us: No matter how fastidious we might be, there is always a danger that the game parameters or the procedures we chose would in some way be biased by our prior beliefs. In fact, most of this data was collected prior to a time when a project such as ours could be envisioned in any more than a cursory way.

The sequential bargaining games we examine are all finite, with monetary payoffs, discount factors, and the number of rounds all commonly known to the bargainers. There are two bargainers, α and β, looking to come to agreement on the split of a bargaining pie, *c*. In the first round, α proposes a split, summarized by σ, the proportion of the pie he proposes to keep (offering 1-σ to β). If β accepts, the pie is divided accordingly. If β rejects, the pie shrinks by discount factors ($\delta_\alpha,\delta_\beta$). The game then proceeds to the second round where roles are reversed, with β making the proposal. The game proceeds in this fashion until the last round; if no agreement is reached, the game ends with both players receiving nothing. The ultimatum game is the one round version of this game. The subgame perfect equilibrium for these games is found by straightforward application of backward induction. The conventional version of this solution assumes a person's preferences are strictly monotonic in own pecuniary payoff. We refer to this as the *pecuniary equilibrium* to keep it straight from the ERC equilibrium, which also employs subgame perfection.

2.1 Sources of the data

Table 1 provides a synopsis of all of the studies that we will investigate in this paper.

The ultimatum game arguably gives the cleanest read social preference since this game makes the fewest cognitive demands on bargainers (ex., the necessary backward induction is simplest). The Roth et al. (1991) ultimatum game experiment involved 270 participants from 4

Experiment	Initials	#	Pie size	Rounds	Discount factors	Subjects	Times played	Observ.	First offer	Pecuniary Eq. predict.	Rejection rate		Disadvantag. counteroffers		
Roth, Prasnikar, Okuno-Fujiwara and Zamir (1991)	RPOZ	1	$10 or $30	1	n/a	270	10	1350	0.407	0.001	0.264	*(365/1350)*	n/a		a, b
Binmore, Shaked and Sutton (1985)	BSS	1	100 pence	2	(.25, .25)	163	1	81	0.416	0.250	0.148	*(12/81)*	0.750	*(9/12)*	c
Güth and Tietz (1988)	GT	1	5 to 35 DM	2	(.10, .10)	42	1	21	0.281	0.100	0.190	*(4/21)*	0.750	*(3/4)*	d
		2	5 to 35 DM	2	(.90, .90)	42	1	21	0.427	0.900	0.619	*(13/21)*	0.000	*(0/13)*	
Neelin, Sonnenschein and Spiegel (1988)	NSS	1	$5	2	(.25, .25)	80	1	40	0.274	0.250	0.225	*(9/40)*	0.556	*(5/9)*	e
		2	$5	3	(.50, .50)	80	1	40	0.472	0.250	0.050	*(2/40)*	0.500	*(1/2)*	
		3	$5	5	(.34, .34)	80	1	40	0.342	0.250	0.125	*(5/40)*	0.400	*(2/5)*	
		4	$15	5	(.34, .34)	30	4	60	0.359	0.250	0.156	*(7/60)*	0.857	*(6/7)*	
Ochs and Roth (1989)	OR	1	$30	2	(.40, .40)	20	10	100	0.413	0.400	0.100	*(10/100)*	0.600	*(5/10)*	
		2	$30	2	(.60, .40)	20	10	100	0.487	0.400	0.150	*(15/100)*	1.000	*(15/15)*	
		3	$30	2	(.60, .60)	16	10	80	0.473	0.600	0.188	*(15/80)*	0.733	*(11/15)*	
		4	$30	2	(.40, .60)	20	10	100	0.457	0.600	0.200	*(20/100)*	0.550	*(11/20)*	
		5	$30	3	(.40, .40)	20	10	100	0.433	0.240	0.120	*(12/100)*	1.000	*(12/12)*	
		6	$30	3	(.60, .40)	20	10	100	0.447	0.160	0.140	*(14/100)*	0.857	*(12/14)*	
		7	$30	3	(.60, .60)	18	10	90	0.453	0.235	0.144	*(13/90)*	0.462	*(6/13)*	
		8	$30	3	(.40, .60)	18	10	90	0.467	0.350	0.289	*(26/90)*	0.885	*(23/26)*	
Bolton (1991)	B	1	$12	2	(.67, .33)	16	8	64	0.400	0.333	0.188	*(12/64)*	0.833	*(10/12)*	
		2	$12	2	(.33, .67)	14	7	49	0.482	0.666	0.184	*(9/49)*	0.200	*(2/9)*	
		3	$12	Trunc.	(.67, .33)	16	8	64	0.407	0.333	0.391	*(25/64)*	0.960	*(24/25)*	f
		4	$12	Trunc.	(.33, .67)	16	8	64	0.653	0.666	0.266	*(17/64)*	0.000	*(0/17)*	
Güth and van Damme (1998)	GvD	y	DG 24	3-person	n/a	36	6	72	0.276	0.042	0.097	*(7/72)*	n/a		b, g
									0.065	0.042					

(a) Numbers reported are aggregations of four treatment run, respectively, in Israel, Japan, Slovenia and the United States. Payoffs in local currency; size of pie outside of U.S. so that "purchasing power on the high side of $10."

(b) In these games, rejections led automatically to disagreement.

(c) Data reported for Game A. Game B of the experiment solicited first offers but was not actually played, and hence is not reported.

(d) The 42 subjects played both games, reversing roles in between. Pie sizes and discount factors were assigned at random across the two games. In this study, a disadvantageous counteroffer automatically led to the disagreement outcome.

(e) The same 80 subjects participated in the first three games.

(f) For the truncation games, the second period responder was restricted to accepting the offer.

(g) Top number refers to mean offer to the responder, and bottom mean offer to the dummy. Minimum offer allowed: 5 tokens to each player (out of 120).

Table 1 – Summary of the experimental designs and observations – average first offers, rejection rates and disadvantageous counteroffers – for each bargaining study in the sample.

different countries (Israel, Japan, Slovenia and the U.S.) who each played a succession of 10 ultimatum games, for a total of 1,350 observations. To control for differences in exchange rates, the U.S. treatments included one with a bargaining pie of $10 as well as one of $30. The study compared ultimatum bargaining play to that in a simple market game with a similar pecuniary equilibrium path. With regard to the ultimatum game, across the sample, offers averaged 40.7% with 26.4% of the offers rejected. There was no significant difference in these figures with respect to pie size. The experimenters identified modest differences among bargaining outcomes across countries, which they attributed to differences in expectations of what would be considered an acceptable offer, rather than differences in propensities "to trespass on a shared notion of what constitutes such an offer" (Roth, 1995, p.288). While all play remained far from equilibrium, there was a notable experience effect.

These results are comparable with those from other ultimatum game studies. Beyond this, the Roth et al. ultimatum game study has four characteristics which, when taken together, make it singular for our purposes. First, bargainers played repeatedly, partners rotated. Several of the sequential offer studies have a similar design (Table 1), and report experience effects that we need have estimates for. Second, the two bargaining pie sizes used in the Roth et al. permit us to fit the "amount of noise" in bargainer decisions, also an important attribute of our model (detailed in section 3). Third, Roth et al. is one of the largest ultimatum game studies, and importantly, includes many observations *per round*. Our model is fit using maximum likelihood estimation, and the desirable asymptotic properties "are justified only in large sample situations" (Eliason, 1993). Fourth, the Roth et al. data was gathered over four countries. Data for the multiple round games comes from three countries (Germany, Great Britain and the U.S.). While our model does not control for cultural differences, the hope is that fitting the model from a multi-country data set mitigates cultural effects from any one country.

The rest of the studies listed in Table 1 are multiple round sequential bargaining studies, save the Güth and van Damme (1998) study, which we will use as a further robustness check of the model (section 5.3). While each of the multiple round studies was performed by a different group of experimenters, they can nevertheless be viewed as a sequence, in that, beginning with Binmore et al. (1985), each study frames itself as a test of the robustness of the results reported in the preceding studies. In fact, the data encompasses a good deal of diversity with respect to game parameters, including pie size, discount factors, and number of rounds, providing a good test case for out-of-sample fits.

(Later studies tend to modify the game or experimental procedures in ways that go beyond the scope of our model. Some sequence or bundle several different games for the expressed purpose of modifying the pattern of subject learning by permitting them to draw inferences across games.[3] The simple quantal response component of our model is unlikely to capture this sort of nuance. For the same reason we do not attempt to investigate the many studies that modify the finite, sequential bargaining game's structure or employ robots in the bargainer pool.[4] Nor do we include papers that replicate results already in our study or were unpublished at the time of this writing.[5])

It will be useful to have a brief synopsis of each study in our sample:

Binmore et al. (1985) performed their experiment on a 2-round sequential bargaining game with a bargaining pie of 100 pence and discount factors (δ_α,δ_β) = (.25,.25). The pecuniary equilibrium first offer is 25 pence (±1 pence since all offers must be in integers). In fact, the average first offer was substantially higher, 42 pence. In a second part of the experiment, those in the β-role were invited to state what they would offer in the α-role. While the game was not actually played to conclusion, (so we do not include it as part of our data set), the responses moved substantially towards pecuniary equilibrium; Binmore et al. put a good deal of weight on this in their conclusions, which also frame one of the critical issues in this line of research. "*The results... indicate little support for the view that a substantial proportion of the population are "fairmen" as opposed "gamesmen"*. (p.1179, italics original).

Güth and Tietz (1988) hypothesized that Binmore et al.'s results were due to a pecuniary equilibrium outcome that intersected with offers that were 'socially acceptable'. They then performed an experiment on a 2-round game, with varying pie size, using discount factors of

---

[3] Ex., Binmore et al. (2002), Carpenter (2003), Harrison and McCabe, (1992).

[4] Alterations include introducing new actions (ex., Binmore, Morgan, Shaked and Sutton, 1991, Rapoport, Weg and Felsenthal, 1990. Zwick, Rapoport and Howard, 1992) or incomplete information (ex., Forsythe, Kennan and Sopher, 1991, Kagel, Kim and Moser, 1996, Mitzkewitz and Nagel, 1993) or linking payoffs across games (ex., Bolton's, 1991, tournaments).

[5] As part of a novel design that recorded the information about pie shrinkage that subjects looked at, Johnson et al. (2002) examined the same three round sequential bargaining game used by Neelin et al. (1988; included in our sample). Johnson et al. stress that the results of their experiment "replicate the stylized facts about offers, rejections, and counteroffers observed in other experiments." In fact, the results are strikingly close to Neelin et al.'s and so we omit Johnson et al. here. Goeree and Holt (2003) is new and unpublished at the time of this writing.

(.1,.1) and (.9,.9), implying pecuniary equilibrium offers of 10% and 90%, respectively. Observed average offers were far from pecuniary equilibrium (shown in Figure 1 below). As with Binmore et al., a considerable number of opening offers were rejected (Table 1). Güth and Tietz also observed a marked tendency to bargain longer when time costs were lower: rejection rates were dramatically higher in the .9 condition (62%) than in the .1 condition (19%). They also report a small decrease in offers as pie size increased.

Neelin, Sonnenschein and Spiegel (1988) responded to Binmore et al. in a different way, studying 3 and 5, as well as 2, round games. In all of their games, and like Binmore et al., the pecuniary equilibrium opening offer was 25%. In the first experiment, each game was played once, with a $5 pie. In a second experiment, the 5-round game was played with a $15 pie and repeated 4 times (Table 1). In contrast to Binmore et al, the observed average opening offer in the 2-round game was close to pecuniary equilibrium (27%). Offers in the 3 and 5-round games, however, were further away (47% for the 3-round, 34% for the $5 5-round, and 43% for the $15 5-round game). The authors say, "[I]t is natural to consider the hypothesis that [first proposers] act myopically: They always act as if they were in the two-round game." p.827. As with Binmore et al., a considerable number of first offers were rejected. Neelin et al. found no evidence that offers vary with pie size. In the end, they reject both pecuniary equilibrium as well as the "equal-split model" as an explanation for their results. In a reply to Neelin et al., Binmore et al. (1988) noted that the differences across studies are plausibly due to differences in experimental procedures.

Ochs and Roth (1989) tested 2 and 3-round bargaining games. The games involved a pie of $30 divided into 100 tokens; the latter feature permitted them to study games with asymmetric discount factors (Table 1). This is the biggest multiple round study in our sample, with a total of 8 conditions (2-round or 3-round, 4 combinations of discount factors per version). Each game was repeated 10 times. Ochs and Roth identify a series of regularities in their data, including a consistent first mover advantage in that the average opening offer favors α even when his pecuniary equilibrium share is less than half; the α-player's discount factor influences the bargaining even in the two-round game where pecuniary equilibrium suggests that only the β-player discount factor should matter; observed mean offers tended to deviate from pecuniary equilibrium in the direction of equal division; and a substantial percentage of rejected offers are followed by 'disadvantageous counteroffers'; that is, counteroffers that give the proposer less than his or her share of the turned down proposal.

Bolton (1991) studied two-round sequential bargaining games with two different sets of discount factors ($^2/_3$, $^1/_3$) and ($^1/_3$, $^2/_3$). The pie size was \$12 and the games were repeated 7-8 times. One of the novel features of this experiment was that it studied what happened when the second round responder was restricted to accepting any second round offer (the truncation games). The ($^1/_3$, $^2/_3$) version of this game provides one of the few instances in multiple round bargaining games where the first mover advantage fails (see section 0), and so provides an important out-of-sample challenge for our model. (Bolton, 1991, also studied games in which bargainers were paid in tournament style. Attempting to fit these would require quite substantial alteration of our model, and so we do not include them in the sample.[6])

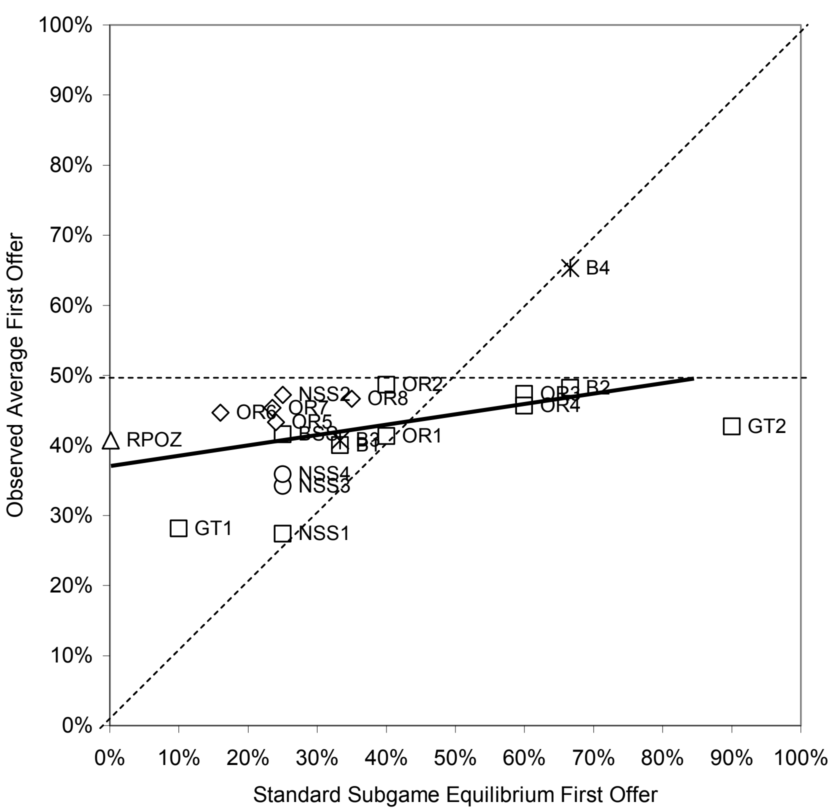


Figure 1 – A comparison of observed average opening offers to pecuniary equilibrium opening offers. The serrated 45° line indicates where prediction and observation match. The horizontal serrated line marks equal division offers. The solid line is the regression line (numbers in parentheses are two-sided p-values).

Regression line:

$$\text{Observed} = \underset{(.110)}{0.14}\ \text{Predicted} + \underset{(.000)}{0.38}$$

$$R^2 = 0.144$$

Legend:

△ Ultimatum game

□ Two-round bargaining

◇ Three-round bargaining

○ Five-round bargaining

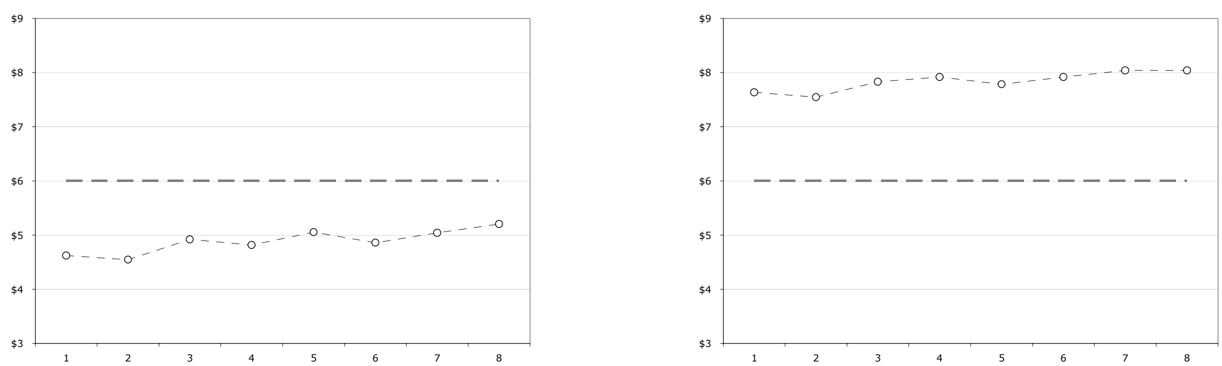
✳ Two-round truncation

---

[6] The tournament games are, because of the payoff alterations made, repeated games between all players in the bargaining session. Applying a model of mistakes requires consideration of bargainer beliefs about the play of others off-the-equilibrium path, a substantial complication. We also do not include two other treatments, both of which yield results very similar to those we do consider, one in which roles were alternated and another with more experienced bargainers.

An overview of the dataFigure 1 shows how the observed average opening offers for these studies compare to pecuniary equilibrium prediction (the labels in the figure refer to the labels in Table 1; Roth et al. is also included for comparison purposes). Altogether, observed average opening offers for multi-round games vary over a fair range (about 27% to 67%), although not by as much as predicted (10% to 90%). Regressing observed average opening offers on those predicted by pecuniary equilibrium yields the results displayed in Figure 1. We see that the regression is not quite significant, with pecuniary equilibrium explaining only 14.4% of the variance in observed opening offers. Moreover, the hypothesis that the coefficient of *Predicted Average Offer* is equal to 1 is easily rejected (one-tailed p<.001). The bias towards higher than predicted offers is evident in the intercept term of .38.

Table 1 summarizes rejection and disadvantageous counteroffer behavior. There is a good deal of across-study variability in first offer rejection behavior, which we seek to fit with our model. Outside of social utility models, it is difficult to provide a theoretical benchmark for this behavior. Pecuniary equilibrium, for example, predicts no rejections. Reinforcement learning presupposes rejection behavior but does not provide an account of how that behavior relates to the parameters of the game.



$(\delta_\alpha, \delta_\beta) = (2/3, 1/3)$ $(\delta_\alpha, \delta_\beta) = (1/3, 2/3)$

Figure 2 – Mean opening offers in the truncation version of sequential bargaining, with a bargaining pie of $12 (Source: *Bolton, 1991*).

Figure 2 illustrates the subtleties in the experience trends observed for opening offers using data from Bolton (1991). For the games in the column to the left, pecuniary equilibrium calls for an opening offer of $4; whereas for the games in the column to the right, it is $8.

Observe that, for the games on the left, opening offers tend to move away from pecuniary equilibrium and towards 50-50 ($6). This is as opposed to games on the right, where opening offers move strongly towards pecuniary equilibrium and away from 50-50.

3 THE MODEL AND THE FITTING TO THE ULTIMATUM GAME DATA

We first lay out the model in its component parts: a social utility function and the quantal response framework. We then fit it to the Roth et al. ultimatum game data.

3.1 The social utility function

The social utility function we use is essentially the one suggested by Bolton and Ockenfels (2000; p.173) restricted to the kind of asymmetry suggested by Bolton's (1991) comparative bargaining model. The function is given by Equation 1.

$$U(\sigma) = \begin{cases} c\left(\sigma - \dfrac{b}{2}\left(\sigma - \dfrac{1}{2}\right)^2\right) & \text{if} \quad \sigma < \frac{1}{2} \\ c\sigma & \text{if} \quad \sigma \geq \frac{1}{2} \end{cases}$$

Equation 1 – Utility function used throughout this paper. *c* is the size of the pie, σ the proportion of the pie the player gets, and *b* measures the relative importance of relative gains for negative reciprocity.

Absolute and relative payoffs are additively separable. Relative payoff is characterized as an asymmetric loss function, with minimum loss (0) when the player obtains half or more of the bargaining pie. The change in marginal utility due to the relative component is greater the further the player's share is below half. The function has one fitted parameter, *b*. The value of *b* no doubt varies with individuals. We interpret our estimate as the population average.

It is important to note that the asymmetry of the utility function is a statement about bargaining game data and not the character of the utility function *per se*. Bargaining games are sometimes referred to as games of 'negative reciprocity' since the tendency for bargainers to reject unfair offers largely dominates any tendency to make fair offers in the sense that the former tends to be the marginal influence on proposer offers (perhaps the clearest empirical demonstration of this principle is given by Forsythe et al., 1994). Later we will see that it is impossible to statistically distinguish the completely asymmetric formulation of Equation 1 from symmetric formulations (section 5.2); and it will turn out that the asymmetric formulation makes better out-of-sample forecasts than a strictly symmetric formulation (section 5.3).

Also note that, in our formulation, the size of the bargaining pie, *c*, does not affect the weights a bargainer gives to relative versus absolute payoffs. That is, players are driven by absolute and relative payoffs in the same proportion across pie sizes, although, importantly, a player prefers a given share from a larger pie than a smaller one. Nevertheless, in this model, pie size does influence decisions, as we will see next.

3.2 The decision making framework: Quantal response

For the sake of exposition, we cast the discussion in terms of the ultimatum game; extension to multiple round sequential offer games is straightforward.

3.2.1 The responder's decision

Let $P_\beta(\sigma_i)$ be the cumulative probability that the responder, β, accepts an offer of proportion σ_i of the pie. By definition, $P_\beta(\sigma_i) \in [0,1]$ for all *i*. $P_\beta(\sigma_i)$, is expressed by a logit function:

$$P_\beta(\sigma_i) = \frac{e^{\tau_\beta . U(\sigma_i)}}{e^{\tau_\beta . U(\varnothing)} + e^{\tau_\beta . U(\sigma_i)}} = \frac{e^{\tau_\beta . U(\sigma_i)}}{1 + e^{\tau_\beta . U(\sigma_i)}}$$

Equation 2 – Responders' probability to accept an offer of σ_i.

where $U(\varnothing)$ and $U(\sigma_i)$ are the utilities of rejecting or accepting the offer, respectively, and are calculated from Equation 1. Rejecting shrinks the size of the pie to *c*=0, so $U(\varnothing) = 0$ for all σ_i.

We call τ_β>0 the *coefficient of certitude*. In this model, τ_β is an indicator of individuals' choice consistency. As $\tau_\beta \to \infty$, if $\mathrm{U}(\sigma_i) > \mathrm{U}(\varnothing)$, then $\mathrm{P}_\beta(\sigma_i) \to 1$; if $\mathrm{U}(\sigma_i) < \mathrm{U}(\varnothing)$, then $P_\beta(\sigma_i) \to 0$. That is, the larger τ_β, the higher the probability the responder follows the strategy that produces the highest utility. At the other extreme, if τ_β=0, there is a 50-50 chance of accepting an offer, independent of the actual value of σ_i. That is, offers are accepted or rejected arbitrarily. It has been observed in various experiments that players are sometimes inconsistent over time, accepting an offer in one game, and refusing a somewhat better offer in another. The probabilistic nature of the decision rule, introduced when τ_β takes a relatively small value, is consistent with this phenomenon and introduces some uncertainty in the strategy the *same* player will follow over time.

3.2.2 *The proposer's decision*

Let $P_\alpha(\sigma_i)$ be the cumulative probability that the proposer, α, makes an offer of σ_i. By definition, $\sum_{i=0}^{I} P_\alpha(\sigma_i) = 1$. Consistent with the responder's decision, the proposer's decision to offer σ_i (and to offers to keep $1-\sigma_i$ for him or herself) follows a logit distribution:

$$P_\alpha(\sigma_i) = \frac{e^{\tau_\alpha . \mathrm{E}(U(1-\sigma_i))}}{\sum_{j=1}^{I} e^{\tau_\alpha . \mathrm{E}(U(1-\sigma_j))}} = \frac{e^{\tau_\alpha . P_\beta(\sigma_i) . U(1-\sigma_i)}}{\sum_{j=1}^{I} e^{\tau_\alpha . P_\beta(\sigma_j) . U(1-\sigma_j)}}$$

Equation 3 – Proposers' decision rule: probability for the proposers to make an offer of σ_I. Right part: re-expressed as a function of responders' probability to accept such offer

$\mathrm{E}(U(1-\sigma_i))$ is the expected utility of offering σ_i to the responder. Consistent with the perfect Bayesian equilibrium solution concept, we suppose that proposers know the true probability responders will accept any particular offer, so the expected utility of an offer σ_i is equal to $P_\beta(\sigma_i) \cdot U(1-\sigma_i)$.

It follows that $\sum_{i=0}^{I} P_\alpha(\sigma_i) = 1$, and that the offers with the highest expected utilities are likely to be chosen more often. Again, as $\tau_\alpha \to \infty$, proposers systematically make the offer σ_i that procures the highest expected utility. If $\tau_\alpha = 0$, $P_\alpha(\sigma_i) = P_\alpha(\sigma_j)$ for all *i*, *j*.

3.2.3 *Modeling the coefficient of certitude*

We model the coefficient of certitude as

$$\tau_k = (\tau_0 + \tau_1 g) \cdot \ln(n), k = \alpha, \beta$$

Equation 4 – Decision parameters of the players re-expressed as a function of a common parameter τ_0 and an experience trend τ_1, weighted by the number of alternatives. The value *g* is the number of games already played, and *n* is the number of choices available to the player.

A key assumption of multinomial logit is the independence of irrelevant alternatives; that is, expected utilities of each strategy should be independent of one another. In the context of our model, absent some compensation, this assumption would be violated since expected utilities in the responder's case are point estimates of an underlying, continuous utility function. Arbitrarily increasing the number of choices confronting the bargainer, say, by increasing the number of tokens players bargain over, without modifying the size of the pie, would add artificial noise. We compensate by multiplying the decision parameter by the log of the number of options.

We discount the coefficient of certitude at a rate dependent on the amount of experience the bargainer has, so that the decision parameter, and so the amount of decision certainty, increases with experience.

Note also, from equations 1 to 3 that increasing the size of the pie has the same effect in this model as increasing the coefficient of certitude. Consequently, the model structure conveys that players select actions with the highest expected rewards more frequently when the size of the pie is large. In other words, players' decisions are expected to be more consistent, less erratic, when the game's stake is higher.

3.3 *Estimating the parameters of the model from the ultimatum game data*

The model has three estimated parameters: *b*, the unique parameter of the utility function, and τ_0 and τ_1, the two parameters that drive players' coefficients of certitude. The other parameters, *c*, *n*¸ and *g* are characteristics of the game design and so given *a priori*.

We fit the model to the multi-country bargaining experiment conducted by Roth et al. using maximum likelihood estimation. As nothing is assumed about the underlying process that generated the data, standard deviations are estimated using nonparametric bootstrap variance estimation (see Davison and Hinkley 1997), and are shown within parentheses. The parameter estimates we obtained are

b=10.742	τ_0=.3478	τ_1=.0159
(.995)	(.0189)	(.0038)

All parameters have the expected sign, and are significant at $p<.01$. As to the fit, the model predicts that the average offer will be 4.06, with an average rejection rate of 30.9%. The corresponding observations are 4.07 and 26.4%. Student t-tests at $p<.05$ on these two measures cannot reject the model. The correlations between observations and predictions are high, with R_α=.907 and R_β=.973. The fit can also be examined in the following table and figures:

σ_i	$P_\alpha(\sigma_i)$		$P_\beta(\sigma_i)$	
	Observations	Model	Observations	Model
0.0	.006	.006	.000	.026
0.1	.020	.012	.333	.113
0.2	.107	.053	.424	.312
0.3	.076	.223	.534	.561
0.4	.416	.359	.714	.734
0.5	.338	.253	.928	.811
0.6	.025	.083	.855	.847
0.7	.003	.010	1.000	.873
0.8	.002	.000	1.000	.893
0.9	.000	.000	1.000	.909
1.0	.007	.000	1.000	.924

Table 2 – Probability for the proposer to make an offer of σ_i, and probability for the responder to accept such offer: observations vs. model (source: *Roth et al., 1991*).

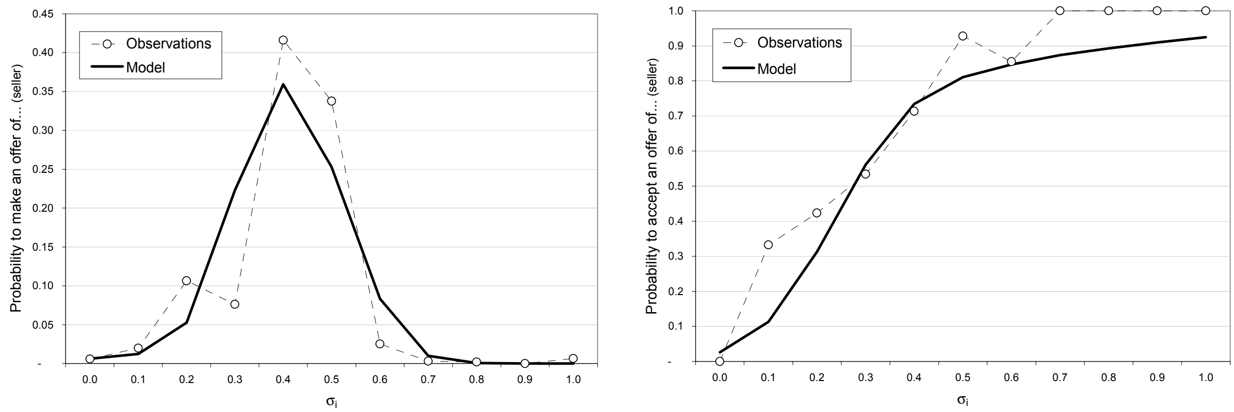


Figure 3 – Probability for the proposer to make an offer of σ_i (left), and probability for the responder to accept such an offer (right), observations vs. model (source: *Roth et al., 1991*).

Recall that we assumed perfect asymmetry (exclusively *negative* reciprocity) with respect to the relative payoff. This is not to say that we have evidence against positive reciprocity. We do not. The log-likelihood of the asymmetric model was -5908, with a standard deviation of 80.1 (bootstrap variance estimate). The log-likelihood of the model with perfectly symmetric relative payoff considerations ($b \neq 0$ for all σ) was -5840, with a standard deviation of 76.2. We cannot reject the null hypothesis that there is no positive reciprocity considerations involved in this game, and that no statistical difference could be found between the two models. We retain the original, asymmetric model. The asymmetric model is simpler, and in section 5, we will show there are reasons to prefer the asymmetric model.

Figure 4 shows how experience affects mean opening offers (model versus observations).

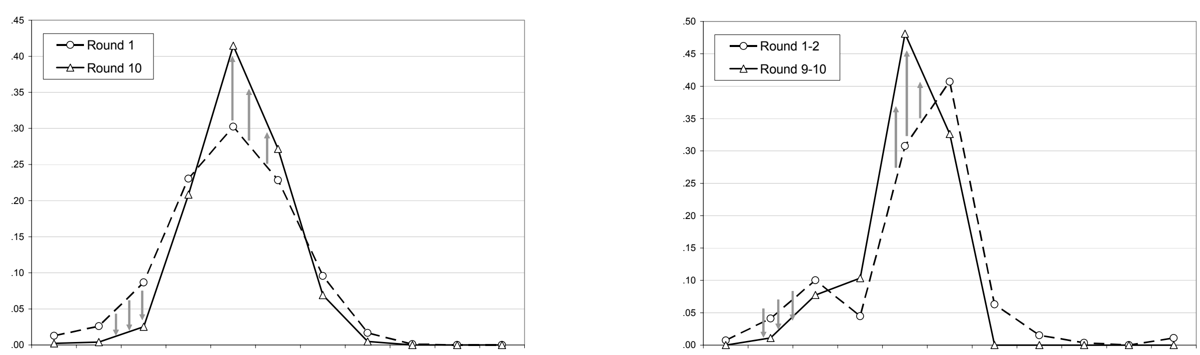


Figure 4 – Probability for the proposer to make an offer of σ_i during the first and last periods, model (left chart) versus observations (right chart). Bargainers find that small offers are likely to be rejected, and eventually make offers that tend to converge around 40% of the pie (source: *Roth et al., 1991*).

4 OUT-OF-SAMPLE FITS TO MULTIPLE ROUND SEQUENTIAL BARGAINING GAMES

We first look at the out-of-sample fits for opening offers and then at rejections and disadvantageous counteroffers. After, we examine how well the model captures experience effects. In all cases, expected utilities are computed using backward induction and supposing that the relevant coefficients of certitude are known and taken into account. The solutions are computed by simulation. Unless specified otherwise, simulations of repeated games are averaged over an equal number of games (with experience trend) as those in the comparison experiment. The Appendix provides a detailed accounting of observed and predicted opening offers, rejections and counteroffers, all broken out by treatment of experiment.

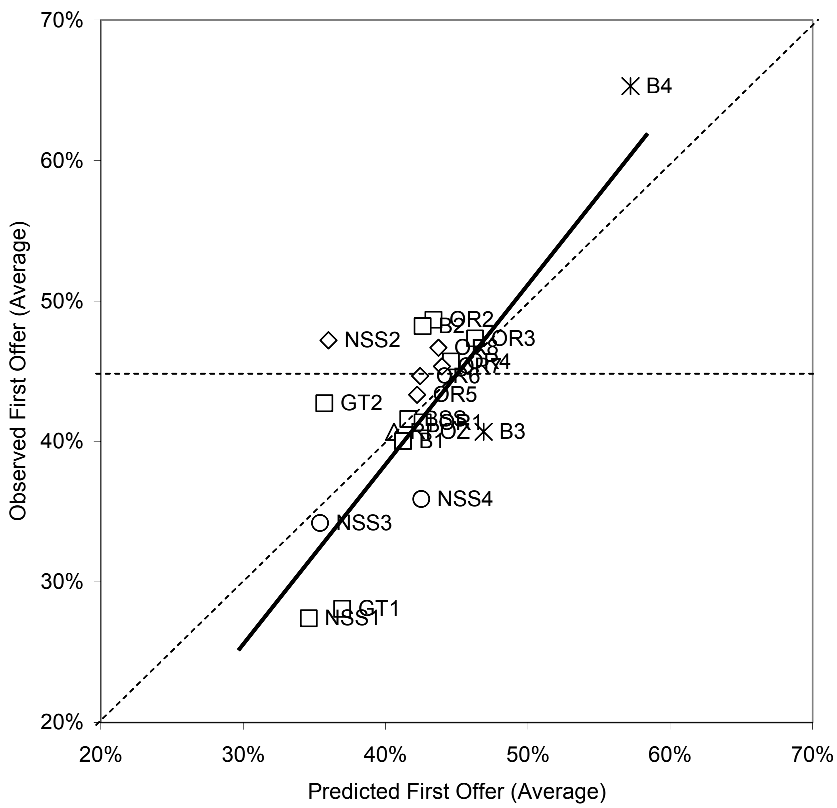


Figure 5 – Average opening offers are predicted quite satisfactorily across experiments and treatments. The solid line in the graph is the regression line. The numbers in brackets are two-sided p-values. If we drop the extreme observation (truncation game), the resulting regression estimate is Observed = 0.97 Predicted + 0.02; R^2=0.333.

Regression line:

$$\text{Observed} = \underset{(.000)}{1.22}\ \text{Predicted} - \underset{(.416)}{0.09}$$

$$R^2 = 0.594$$

Legend:

- □ Two-round bargaining
- ◇ Three-round bargaining
- ○ Five-round bargaining
- ✳ Two-round truncation

4.1 *Opening offers*

Figure 5 plots predicted average opening offers against what was observed for each multiple round study. The figure also provides details for the associated regression estimate. A perfect fit is along the 45° line. We can see the model fits this data quite well. The estimated coefficient for Predicted is 1.22; the model that restricts this value to 1 cannot be rejected (two-tailed p=.561). The small intercept term indicates little in the way of bias in the estimates. We judge the amount of variability explained, R^2 of .594, good given the simplicity of the model. Of course, the model is not as simple as the pecuniary equilibrium model; but the former easily outperforms the latter (compare with results in section 0).

Observe that the observation well outside the range of the rest of the data, the Bolton ($^1/_3$, $^2/_3$) truncation game, is well accounted for. Of course, this observation also accounts for a substantial proportion of the explained variability. Re-estimating with this observation dropped yields an R^2 of .333. The regression coefficient is still close to 1, indicating our results are robust to this observation (see Figure 5 for details).

Looking at Figure 5, there is no apparent bias with regard to individual studies, nor (turning back to Table 1) with respect to pie size. In fact, inserting dummies for all but the Bolton study into the Figure 5 regression, finds none significant at any standard level. This test may have low power due to the sample sizes, but in this regard, all of the estimated dummy variables have low absolute values, none larger than .013. Adding a variable for pie size to the equation in Figure 7 shows no significant effect for pie size at any standard level. These results do not change if we do the analysis without the extreme observation.

Our model was estimated from a game that is itself extreme relative to the out-of-sample games, in two respects. For one, the first mover advantage is extreme, in the sense that an ultimatum game responder rejection leads to the loss of the entire pie (compare to GT2 where rejection leads to only a 10% loss for both bargainers). The most straightforward measure of the first mover advantage, in this sense, is simply the pecuniary equilibrium (note that this measure is *negatively* correlated with first mover advantage). In fact, when we insert the pecuniary equilibrium into the Figure 5 regression as a variable, the estimated coefficient is .13 and it is significant (p=.043). Hence the model has a tendency to underestimate the opening offer as the first mover advantage diminishes. But, in light the relative extremity of the game used for fitting, we judge this bias small.

The second way in which the ultimatum game is extreme is that it is one round. One round puts modest demands on peoples' ability to backward induct. The question then is whether the accuracy of the model falls off as the number of rounds of the game increases. Seven of our 19 games have more than 2-rounds (five of these are 3-round and two are 5-round). We test for a falloff in estimation accuracy by adding a dummy variable (1=more than 2-rounds) to the regression estimated in Figure 5 (all observations). The estimated coefficient is just .02 and is insignificant (*p*=.511). Also, the average size of residual for games with more than 2 round games, .033, is just a bit smaller than it is for 2-round games, .041, so estimates for more than two round games are not less accurate. Inserting separate dummies for 3 and 5-round games yields very similar results.

In other words, there is no evidence here for a loss of accuracy for the higher round games. This should not be interpreted as implying that people “are doing” backward induction in an unlimited sense. There is a convincing body of work showing that people are limited. What the analysis does imply, however, is that the limitation is not so extreme as to prevent backward induction from providing a reasonably good approximation for opening offers, at least where the number of rounds is modest. Also, in section 5.1 we will see that there is a modest but notable difference in parameter estimates when we fit the model using 3-round data, and in this sense there is a difference.

Our model captures several regularities that were noted by the authors of the sample studies. From Figure 1, we see, as Ochs and Roth noted, that opening offers tend to be higher than the pecuniary equilibrium but less than half the pie. The one exception, also captured by the model, is a truncation game, where the offer is considerably higher. Ochs and Roth also noted a tendency for the first mover’s discount factor to affect the outcome (fixing the other discount factor), even in the two-round game. From the appendix, the reader can verify that, for all relevant comparisons, model predicted average opening offers move in the same direction as the data. We also noted earlier that Neelin et al. observed higher first offers in their 2-round game than did Binmore et al., even though the two games had the same discount factors and therefore the same pecuniary equilibrium. One explanation is that the procedures involved in the games differed (ex., different directions, subject pools). A second explanation is suggested by our model; it predicts higher opening offers in the Binmore et al. games (.416 versus .346 although the difference is not quite significant for the two-tailed *z*-test, *p*=.122). The difference, our model suggests, is driven by the differing bargaining pie sizes. In the larger pie game (Neelin et al.), decisions have a smaller arbitrary component, and this is reflected in the accompanying lower proportion offer.

4.2 Rejection rates

Figure 6 displays the out-of sample plots for rejections of opening offers. Again, a perfect fit is along the 45° line. We observe two kinds of deviation from prediction. First, there is a good deal of heteroscedasticity; variability increases as the predicted rejection increases. Second, for the 4 games with the smallest bargaining pies, predicted rejection rates hover around arbitrary, or 50%, substantially higher than observed. The regression that estimates observation from predictions, using weighted least squares, is insignificant and accounts for only 7% of the data. But it is apparent from the diagram that this is due to the bad small pie predictions.

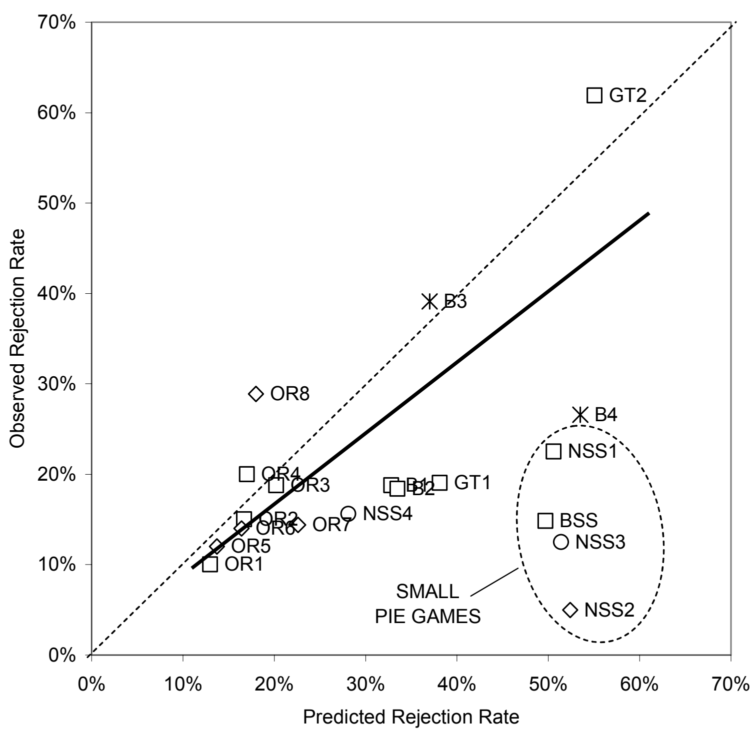


Figure 6 – Regressing observed rejection rates on predicted rates shows little fit because the predictions for the smallest pie games have little value. A regression that corrects for this fits the data fairly well (solid line in the graph). The regression estimates shown are for weighted least squares, to correct for the heteroscedasticity, where the weights are the predicted rejection rates.

Regression line:

$$\text{Obs.} = \underset{(.945)}{0.00} + \underset{(.004)}{0.76}\ \text{Pred.} - \underset{(.001)}{0.25}\ \text{Small}$$

$$R^2 = 0.539$$

Legend:

□ Two-round bargaining

◇ Three-round bargaining

○ Five-round bargaining

✳ Two-round truncation

The weighted least squares regression estimate displayed in Figure 6 adds a dummy variable for the four smallest pie games. Now the fit is much improved, R^2 of .539. The coefficient for PREDICTION is .76. We cannot reject the model that restricts this value to 1 (two-sided p =.396). The intercept term is small and not significant. Some notes are in order:

1. Dropping the small pie games from the sample, and rerunning the regression yields very similar estimates.

2. One might think that, generally, the smaller the pies size, the more likely the model over-predicts. But regression analysis does not bear this out. Neither pie size nor reciprocal pie size variables show any significant correlation with observed rejection rates, nor does either variable add to the predictive value of the Figure 6 model.
3. The predicted rejection rates were derived using predicted offers. There is no evidence, however, that this biases the rejection predictions; specifically, there is statistical correlation between rejection prediction errors and offer prediction errors, nor between there absolute values, at any standard level of significance.
4. The smallest pie games are also the once repeated games, raising the issue of whether single play might not be an explanation for their outlying. Experiments with the simulation routine that produces the predictions suggests there is some truth in this (predicted rejection rates show some tendency to decline with experience), but also implies that repetition is far less important than pie size.

The presumption, built into the model, is that smaller pie size leads to more arbitrary behavior. We have seen that the presumption fits opening offer behavior well. But for rejection behavior, where not coincidentally the amount the responder has at stake tends to be smaller than what the proposer has at stake, behavior is steadier for very small pies than the nearly perfectly arbitrary behavior predicted.

Adding the appropriate dummies to the regression displayed in Figure 6, we find no evidence that rejection rate predictions are, *on average*, better for some individual studies than others. The sample sizes tend to be small, and descriptively, it appears that predictions are more uniformly accurate for the Ochs and Roth study than for the others; but that said, it is difficult to know what to make of it since rejection rates for this study were also less variable across treatments than for the other studies.

Rejection rates predictions show no bias with regard to number of rounds at any standard level of significance. Adding the subgame perfect offer to the regression shows the same sort of bias, with respect to first mover advantage, that we found for opening offers although this time the coefficient, .25, is only weakly significant (two sided p=.092).[7] So again, there is some

[7] Actually, the subgame perfect predictions proved highly collinear with the two variable in the Figure 6 regression. The reported results are taken from an estimate that uses the residuals obtained from regression subgame perfect predictions on our model's predictions and the smallest pie dummy.

modest evidence that our estimates are biased as we move away from the rather extreme case we used to fit our model.

Güth and Tietz observed a dramatic rise in rejection rates when the cost of disagreement was lowered. Our model suggests that this is generally so. Ochs and Roth's experiment includes ten comparisons (holding the number of rounds fixed) where the cost of disagreement increases (i.e., on discount factor decreases and the other stays fixed, or both decrease). In all cases, our out-of-sample estimates suggest rejection rates should rise. In fact, rejection rates rise in eight of the ten cases (p=.055, one-tailed sign test).

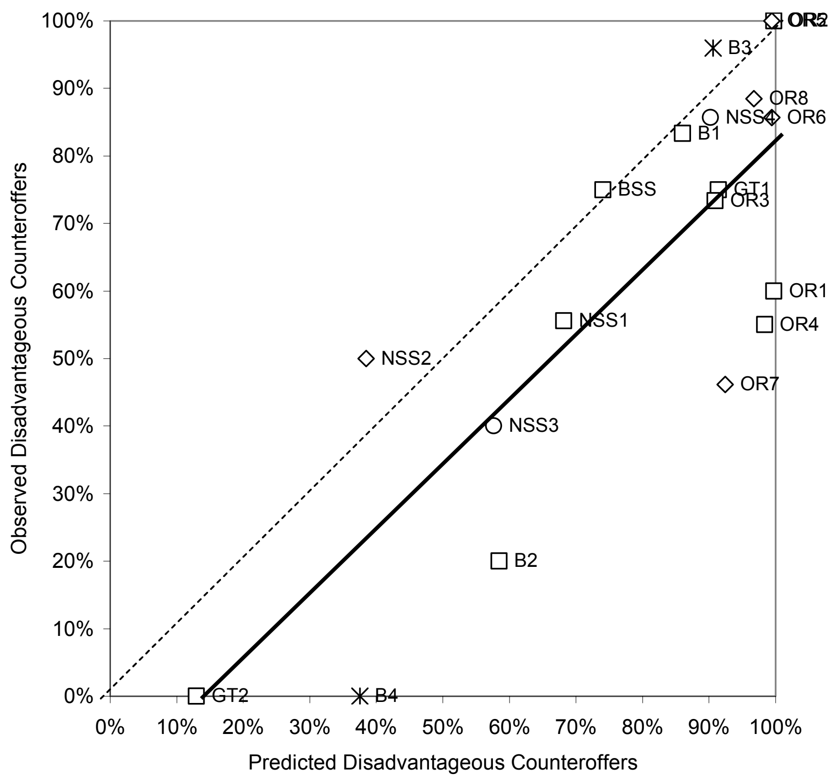


Figure 7 – Proportion of disadvantageous counteroffers is predicted satisfactorily across experiments and treatments, even though this proportion is observed on few observations. There is a modest downward bias to prediction, although this is not statistically significant. The solid line in the graph is the regression line.

Regression line:

$$\underset{}{\text{Observed}} = \underset{(.000)}{0.987}\ \text{Predicted} - \underset{(.307)}{0.144}$$

$R^2 = 0.673$

Legend:

□ Two-round bargaining
◇ Three-round bargaining
○ Five-round bargaining
⁎ Two-round truncation

4.3 Disadvantageous counteroffers

We also estimated the incidence of disadvantageous counteroffers, that is, when responders reject an initial offer to eventually make a counteroffer that leads to a lower monetary payoff. Despite the fact that datasets under scrutiny have little counteroffer behavior (between 2 and 26 per treatment, with an average of 12.4), and hence proportions are computed on few observations, the predictions fit quite well (see Figure 7), with an R^2 of .673.

4.4 Experience effects

Three studies (Neelin et al., Ochs and Roth, Bolton) gave players repeated experience bargaining in the same role (Güth and Tietz had their subjects negotiate once in each role with potentially different discount factors and pie sizes, and for this reason we treat their experiment as single play).

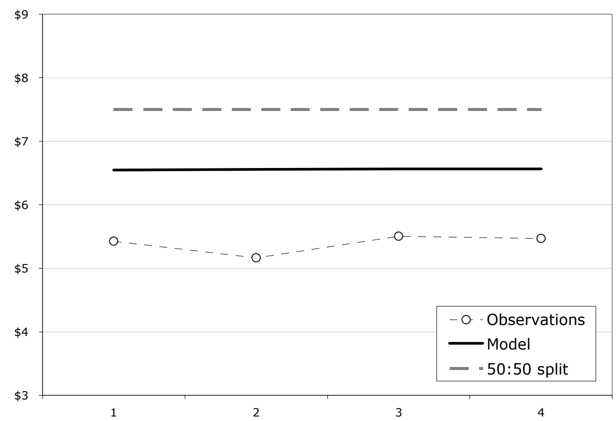


Figure 8 – Mean opening offers by round in a 5-round game, observations versus model (source: *Neelin et al., 1988*).

Neelin et al., in their $15 5-round game, had subjects play 4 rounds (although the first round did not count for payment). Figure 8 shows there is little in the way of an experience effect in the data. The model suggests the same, although the average offer is somewhat overestimated.

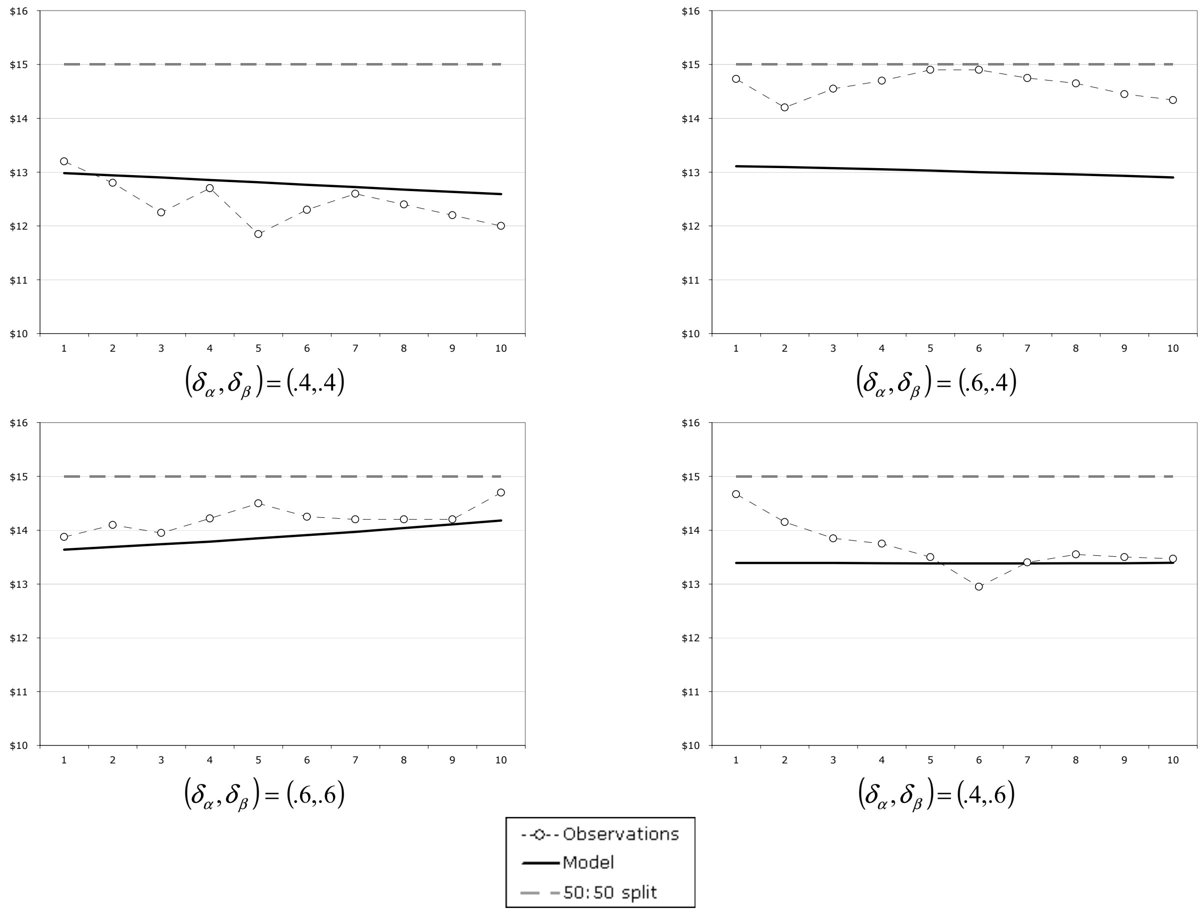


Figure 9 – Mean opening offers by round in 2-round games with different discount factors, observations versus model (source: *Ochs and Roth, 1989*).

Our model also captures most of the experience trends Ochs and Roth observed. Figure 9 compares model and data for the 2-round games. In the (.4,.4) treatment, first players *decrease* their opening offers over time, from 13.2 in the first game to 12.0 in the last game. Our predictions are 13.0 and 12.6 respectively. At the other extreme, in the (.6, .6) treatment of the 2-round ultimatum game, first players *increase* the offer they make to the second player, from 13.9 in the first game to 14.7 in the tenth game, on average. Our model predicts this upward trend, too, from 13.6 to 14.2 after 10 games. The model also predicts the direction of the experience trend in the (.6, .4) treatment, although it consistently underestimates the mean opening average. The model fails to replicate the experience trend in the (.4, .6) treatment during the first few games, although both model and observations eventually converge.

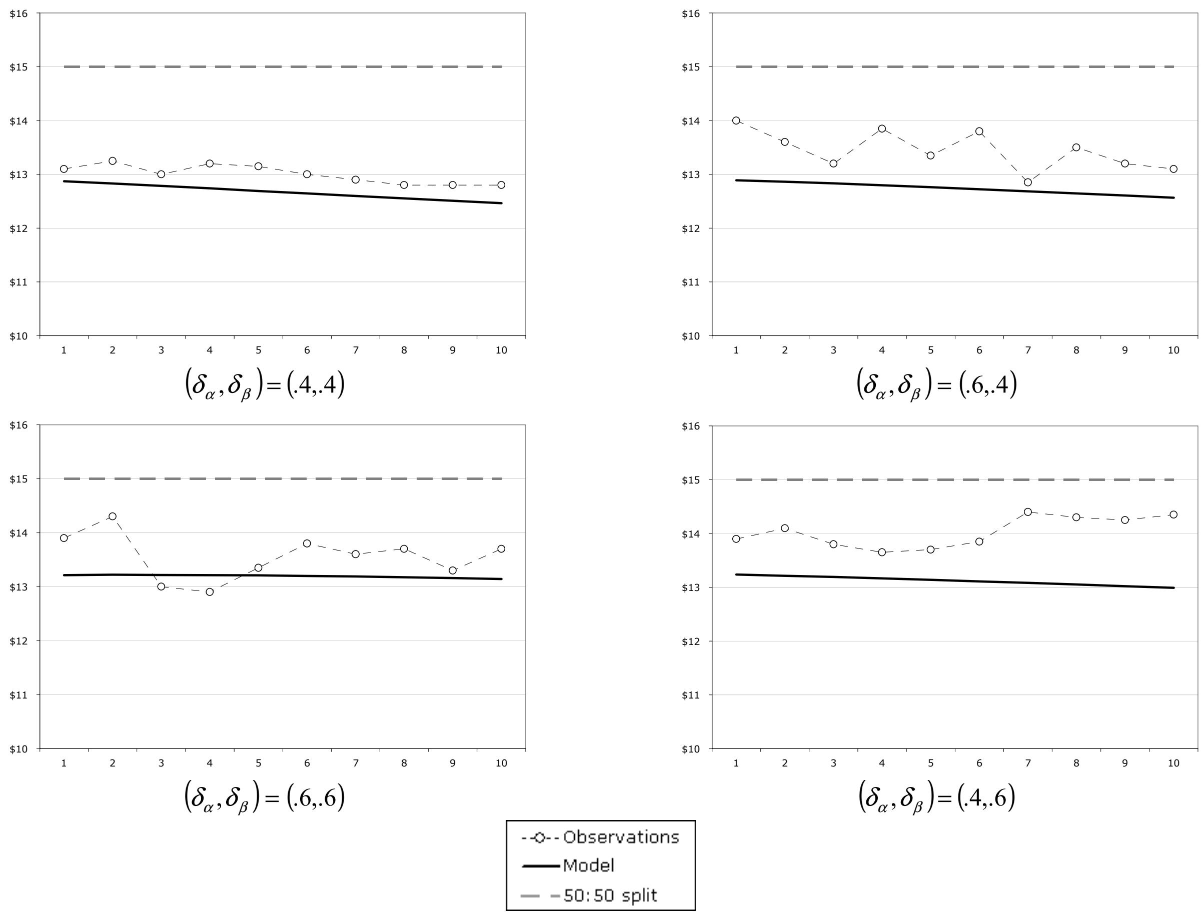


Figure 10 – Mean opening offers by round in 3-round games with different discount factors, observations versus model (source: *Ochs and Roth, 1989*).

Figure 10 displays analogous comparisons for Ochs and Roth's 3-round games. The observed experience trends are quite satisfactorily captured for all treatments except (.4, .6) where both amplitude and direction are off somewhat.

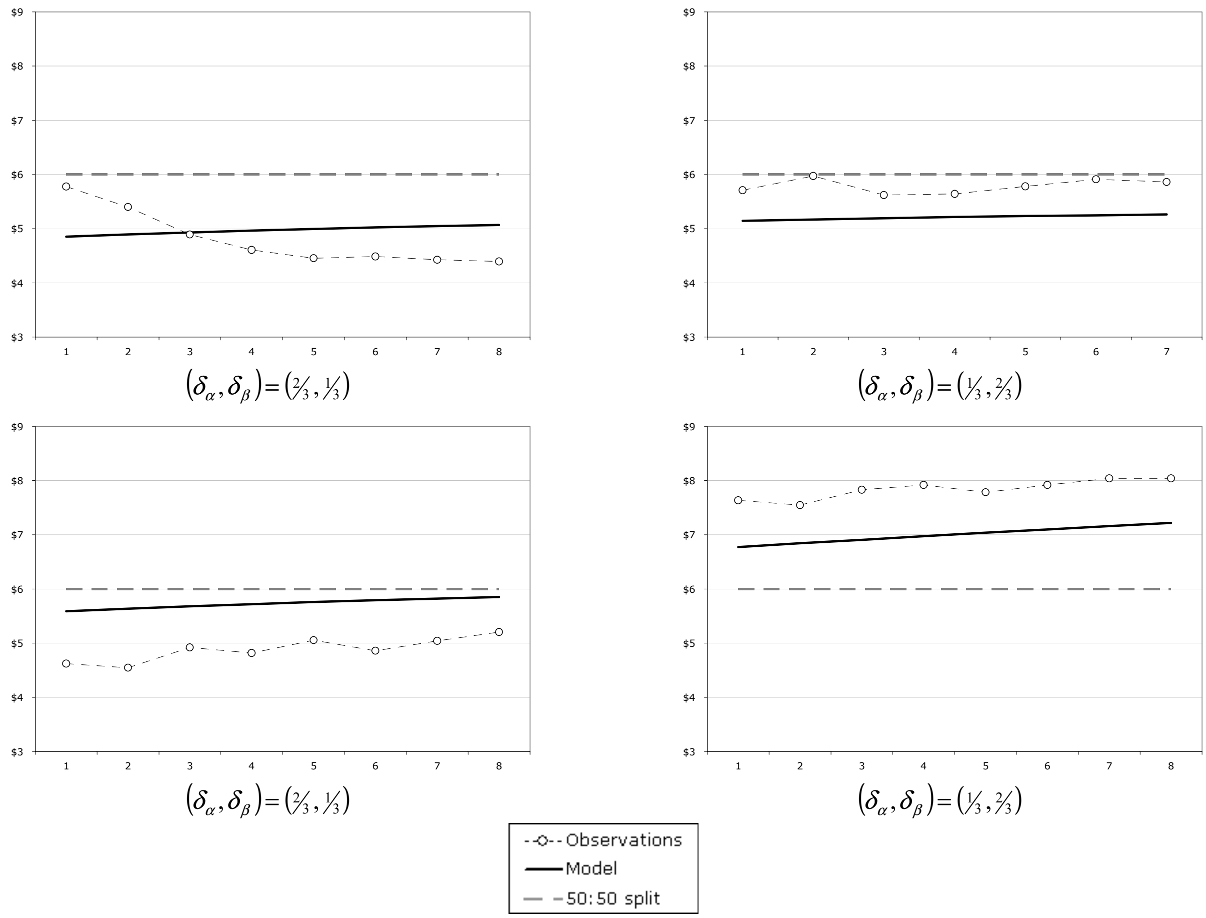


Figure 11 – Mean opening offers by round in 2-round (top figures) and truncation (bottom) version of sequential bargaining, with different discount factors and a bargaining pie of $12 (source: *Bolton, 1991*).

Figure 11 shows the corresponding learning trends for treatments in Bolton (1991). Most noticeably in the $({}^{2}/_{3}, {}^{1}/_{3})$ treatment of the truncation game, observed mean opening offers deviate from the pecuniary equilibrium in the direction of the equal money division. The difference *widens* with experience. In contrast, the difference *narrows* for $({}^{1}/_{3}, {}^{2}/_{3})$. The direction of both of these trends is well captured by the model.

The important conclusion from these graphs is that, with experience, behavior tends to converge to what we would expect from a static ERC model. That is, the model implies that, with experience, the behavior we see is less mistake-prone, and closer to the fully rational ERC equilibrium. By and large, observed behavior moves in the same direction.

4.5 Summary of out-of-sample fits

To summarize our findings regarding the model's out-of-sample fit: The variability in opening offers is well accounted for by the strategic parameters of the game, together with the assumption that offer behavior is more variable for smaller pies (recall the comparison of the Binmore et al. and Neelin et al. 2-round games). There is no apparent bias in our estimates with respect to individual studies. Nor is the fit appreciably different for 3 and 5-round games than for 2-round games.

The model shows little fit with opening offer rejection rates for the smallest bargaining pie games in our sample. However, for the rest of the studies in our sample, it accounts for a good deal of rejection rate variability. There is no apparent directional bias with respect to individual studies, although there is less variability in accuracy with regard to Ochs and Roth's experiment. The model also accounts for the observation that rejection rates rise as time costs unambiguously decrease, holding the number of rounds fixed. In addition, disadvantageous counteroffers are fairly well accounted for by the model, although there does appear to be some downward bias. Finally, the model does a good job of capturing the direction and magnitude of experience effects in most cases move in the direction of the static ERC model.

5 ROBUSTNESS

One issue our estimates raise is how dependent they are on the particular data set we used to estimate the model. In this section, we fit the model in two alternative ways, using the two and three round data from one of the multiple round studies. A second issue is to what extent the key assumptions in our model – an (asymmetric) concern for fairness and the estimated coefficient of certitude – are critical or arbitrary with respect to a successful fit. We check whether a simpler model – one that assumes no decision error, for example – gives notably different results. Finally, we extend our out-of-sample estimates to an experiment on a three-person ultimatum game that the ERC model successfully explains in the ordinal sense (Bolton and Ockenfels, 1998). This provides a strong test of the model's stability with regard to the model's reference point, and also permits a comparison to a data set in which the proposer is far less generous than in any of the games examined so far.

5.1 *Alternative model fits*

To be suitable for estimating the model, a dataset needs to have two critical characteristics: First, to estimate experience effects, the data must include observations of multiple plays of the game. Second, since the desirable properties of maximum likelihood estimates are only achieved asymptotically, the size of the dataset need be rather large. Of the datasets we examine, Ochs and Roth's meets both criteria. In fact, there is enough data here to fit 2-round and 3-round games separately; an interesting exercise given the issue of backward induction's ability to approximate behavior. We therefore re-estimated the parameters of the model using the 2-round and 3-round game data separately. We estimated one set of parameters to fit simultaneously all 4 treatments of each game (i.e., 4 combinations of discount factors), thus leading to two datasets of 380 observations each. To avoid artificial biases, observations were weighted within each dataset so that each treatment would equally contribute to the respective log-likelihood function.

	b	τ_0	τ_1
ULTIMATUM GAME	10.742	0.3478	0.0159
(Roth et al., 1991)	(.995)	(.0189)	(.0038)
TWO-ROUND BARGAINING GAME	10.566	0.2704	0.0016
(Ochs & Roth, 1989)	(2.258)	(.0267)	(.0038)
THREE-ROUND BARGAINING GAME	12.579	0.2206	0.0180
(Ochs & Roth, 1989)	(1.485)	(.0190)	(.0050)

Table 3 – Parameter estimates of the model, using maximum likelihood estimation and 3 different datasets. Estimates seem to be reasonably stable across games.

Table 3 reports the results of the parameter estimates obtained from 3 datasets, the original ultimatum game from Roth et al. used in the first sections of this paper and the 2-round and 3-round versions of the ultimatum game from Ochs and Roth. Most differences are statistically significant, although the *b* estimate is not statistically different between the first and the second datasets, nor is τ_1 between the first and the third datasets. In terms of economic significance, however, parameter estimates appear reasonably similar across games. Perhaps the biggest difference is with respect to the estimate of *b* in the three-round game estimate as opposed to the other two; the three-round estimate implies a somewhat heavier weight on relative payoffs. The heavier weight might be attributable to greater difficulties doing backward induction resulting in decisions that put somewhat more weight on what-is-fair, and less weight on strategic calculation.

5.2 Alternative assumptions about preferences and certitude

In our original estimate, we assumed that the structural form of players' utility function followed an ERC, asymmetric shape (i.e., presence of negative reciprocity, but no positive reciprocity), and that choices were the outcome of a random process. The statistical fit obtained on the simple version of the ultimatum game seemed to confirm our hypotheses, and introducing positive reciprocity did not improve the overall fit of the model. But the null hypothesis that there were no positive reciprocity considerations in players' behaviors could not be rejected.

To gauge how important the asymmetric shape and random process assumptions are to our model, we re-estimated the parameters of the model under all possible combinations of the following assumptions:

1. The decision parameter is set to infinity ($\tau \rightarrow \infty$), transforming the probabilistic decision rule into a deterministic one; the strategy with the highest expected utility would then be chosen with a probability of 1.
2. The parameter *b* of the utility function is set to 0, transforming the nonlinear ERC utility function into a standard linear function; players' motivations become self-interested, monetary gains only, without equity considerations, as suggested by the pecuniary model of classic game theory ("greedy" utility function).
3. The loss function is symmetric, with *b* positive even for $\sigma > ½$, introducing positive reciprocity considerations into players' motivations ("symmetric ERC" utility function).

			Responder					
			Greedy		Asymmetric ERC		Symmetric ERC	
			Deterministic	Probabilistic	Deterministic	Probabilistic	Deterministic	Probabilistic
Proposer	Greedy	Deterministic	-24610 **(a)**	-24557 **(b)**	-24057	-23298	-23928	-24025
			0.001	0.000	0.250	0.300	0.300	0.212
			0.000	0.500	0.000	0.405	0.000	0.420
		Probabilistic	-10864	-6924	-9905	**-5903**	-10012	**-5890**
			0.438	0.336	0.467	**0.408**	0.455	**0.408**
			0.000	0.363	0.047	**0.264**	0.059	**0.275**
	Asymmetric ERC	Deterministic	-24610	-24557	-24057	-21751	-23928	-21910
			0.001	0.000	0.250	0.353	0.300	0.372
			0.000	0.500	0.000	0.264	0.000	0.256
		Probabilistic	-10654	-6685	-9770	**-5908 (c)**	-9892	**-5850**
			0.333	0.308	0.411	**0.406**	0.412	**0.407**
			0.000	0.390	0.093	**0.309**	0.078	**0.331**
	Symmetric ERC	Deterministic	-21644	-20822	-21644	-20822	-21644	-20822
			0.400	0.400	0.400	0.400	0.400	0.400
			0.000	0.199	0.000	0.199	0.000	0.199
		Probabilistic	-10026	**-5862 (d)**	-9643	**-5852**	-9716	**-5840**
			0.432	**0.401**	0.435	**0.407**	0.449	**0.415**
			0.000	**0.225**	0.062	**0.247**	0.052	**0.255**

Table 4 – Log-likelihood, mean opening offer (in percent) and average rejection rate in the simple ultimatum game (c=$10), under different model assumptions. The cells in bold have a log-likelihood not statistically different from our model's.

These changes can be readily applied to the proposer, the responder, or both. Table 4 shows the log-likelihood, the mean opening offer and the rejection rate in the ultimate game after fitting the data using these different model assumptions. Seven of the 36 models tested fit equally well the observations made by Roth et al. (1991), and are boxed and in bold in the table.

For illustration purpose, the cell marked (a) corresponds to the pecuniary equilibrium: The proposer and the responder are both only motivated by their pecuniary gains (greedy utility function, b=0), and they systematically choose the strategy with the highest (expected) utility, hence their choices are deterministic ($\tau\rightarrow\infty$). Therefore, the first player offers the smallest share of the pie possible to the second player, and the latter accepts with a probability of 1. This model is easily rejected by the data.

It has been suggested that, under some treatments, even "greedy" and "rational" players might choose to offer more than the minimum suggested by game theory, in order to secure a rationale response from the second player. For instance, Binmore wrote that "the first player might be dissuaded from making an opening demand at, or close to, the 'optimum' level, because his opponent would then incur a negligible cost in making an 'irrational' rejection" (Binmore et al. 1985, p.1180). In other words, although both players are greedy, the randomness associated with the second player's decision to reject small offers has to be compensated (by increasing the share offered to the second player), so that the cost of an irrational decision of rejection would increase. Cell (b) captures this reasoning, but the hypothesis does not successfully fit the data. On the contrary, our fit suggests that proposers are much better off offering nothing to the second players, and letting them respond randomly with a 50:50 chance of accepting the offer (since responders are greedy, there is no negative utility associated with accepting an unfair offer, and they are hence indifferent between accepting and rejecting an offer of 0).

Cell (c) is the standard model we used throughout this paper: players share the same utility function (an ERC, asymmetric curve), and both players' decisions are probabilistic.

Interestingly, as shown in cells in bold, only two conditions suffice to replicate the observations made by Roth et al. (1991), and all achieve a log-likelihood not statistically different from our original model's. First, responders must have an aversion to games outcomes that are in their disfavor (b_β>0 if σ<½), whether this aversion is embedded in a symmetric or asymmetric utility function; and second, both players' choices must be probabilistic. Characteristics associated with the first player's utility function are irrelevant to explain the results.

As shown additionally in cells (d), proposer's fairness motives could also explain, by themselves, a large mean opening offer. But this hypothesis is refuted by observations made in other games: for instance, in the three-person ultimatum bargaining game which is our next subject.

5.3 A further stress test: Three person ultimatum bargaining

The three-person ultimatum bargaining game is similar to the simple, one-round version of the ultimatum game, except that there is a third player (the "dummy") with whom the pie has to be shared (the player does not actually make any decision). The first player proposes to the second player a division of the pie among all three players, and the second player either accepts or rejects the offer. If the proposition is rejected, all players receive zero monetary payoff. We compare our predictions to the simplest version (i.e., essential information treatment, constant mode) of the original experiment conducted by Güth and van Damme (1998) (Kagel and Wolfe, forthcoming, present similar findings). Players had to share a pie of 24 Dutch Guilders (divided into 120 tokens), which represented by the time approximately \$13.6 (c=13.6).[8]

ERC stipulates a modification in the utility function to fit a 3-person game: Since 3 players are involved, the social reference share of the payoff is one-third instead of one-half of the pie; in terms of the utility function in Equation 1, the term $(\sigma - \frac{1}{2})$ is replaced by $(\sigma - \frac{1}{3})$. Otherwise, we get out-of-sample estimates for this game using the same procedure as before; in particular, the b and τ parameters of the equation are as estimated from the Roth et al. ultimatum game with a social reference share of one-half (Bolton and Ockenfels, 1998, study this game, using the same modification).

One of Güth and van Dammes' critical findings was that the dummy player received little more than the minimum, 5 tokens, that the experimenters required be given. On average during the first six games, the dummy's share was 7.8 out of 120 tokens in the observations (6.5%). Our model predicts 8.4 (7.0%), as shown below.

	Observations	**Model**
Proposer (x)	79.1	76.3
Responder (y)	33.1	35.3
Dummy (z)	7.8	8.4

Table 5 – Average amounts (pie size of 120 tokens, minimum share of 5 tokens per player allowed) allocated to the three players by the proposer in the essential information treatment of the Güth-van Damme game, observations versus model (source: *Güth and van Damme, 1998*).

A second important finding was the rejection rates in this game tended to be smaller than in regular ultimatum games. The average rejection rate in the two-person Roth et al. ultimatum experiment was .264, consistent although a bit higher than the typical 15-20 percent rejection rate observed in 2-person ultimatum games (Roth 1995). It is .097 in the Güth-van Damme original dataset, essential information condition (p.241). Our model does not capture this finding, predicting a rejection rate of .281. Here the problem appears to be that real players learn more quickly than our simulated players do: The predicted rejection rate after 40 games is .094.

[8] Güth and van Damme also report data when the responder knows the entire proposed allocation. This data is very similar to the data we report.

There was also an experience effect in the data, and our model does capture some of its characteristics. We apply the model by predicting the game's outcome at the first, sixth, twentieth and fortieth games. As shown in Table 6, this does not affect *y*'s payoff much, but increases the proposer's payoff to the detriment of the dummy's. In other words, the proposer learns that he can keep the dummy's share of the pie without affecting the responder's likelihood of accepting his proposals. The exact same pattern has been found during Güth and van Damme's actual experiment (p.239).

Model	Round 1	Round 6	Round 20	Round 40
Proposer (x)	75.9	76.7	78.7	81.6
Responder (y)	35.0	35.5	35.2	33.1
Dummy (z)	9.2	7.8	6.1	5.3
Rejection rate	.306	.256	.161	.094

Table 6 – Division of the pie when players gain experience. The model replicates observations: dummy's payoff decreases and proposer's payoff increases with learning.

Again, the model correctly estimates both the nature and the direction of players' learning, but underestimates the pace at which it will occur. Actually, the predictions made by the model for the fortieth game are very close to the observations already made during the sixth game of Güth and van Damme's experiment, namely x=80.8, y=33.3 and z=5.8, with a very similar rejection rate. It seems that the large overestimation of the rejection rate (.281 predicted in early games versus .097 observed) is mainly the consequence of the model's inability to predict the pace at which learning will occur, rather than the direction, nature or effects of such learning.

6 SUMMARY

The model we investigated here is strategic in nature. Players pursue their interest in fairness in a rational way, up to a mistake parameter, taking into account the behavior, and likely behavior, of others. The major conclusion we draw is that a simple model, that modifies the conventional game theoretic model of sequential bargaining by permitting a preference for fairness, and admits some noise in choices, provides quite good estimates of bargainer behavior. The model makes three particular simplifications that appear to work reasonably well:

First, accounting for bargainers' concern to be treated fairly, as opposed to their concern for treating others fairly, is sufficient to get pretty good estimates of bargaining behavior. There is no evidence that adding a concern for treating others fairly would improve the model's fit with the data. This is not to imply that bargainers do not care about treating others fairly; there is plenty of evidence that fair-treatment-for-others is important in other kinds of games. But the tendency for people to care more about being treated fairly than with treating others fairly, when combined with the offer-counteroffer format of sequential bargaining (proposers never offer something unacceptable to themselves) permits the simpler, asymmetric fairness formulation with little loss of accuracy.

Second, look ahead, reason back – backward induction – provides a reasonable fit for all these cases. Both 3 and 5-round out-of-sample predictions are not much less accurate than 2-round predictions. Of course, there is no way to judge from the present data how well the model would fit for a substantially longer game. Nevertheless, we judge the model reasonably stable in this regard.

Third, experience, by and large, moves bargainers towards the equilibria implied by the static ERC model. With experience, our model suggests that choices get more certain and closer to optimal relative to what the ERC model (without random perturbations of choices) predicts.

While the simple model approximates well, it is natural to think about refinements of the model. Here are some issues that such models will have to deal with: The experience component of the ERC utility-decision framework we have used is independent of players' past actions. This seems unrealistic. The experience trend embedded in the model is expected to arise from a decrease in players' heterogeneity and in choice randomness, but the way these phenomena are linked to experience and trials and errors needs to be investigated.

A second issue is the way individual variability is accounted for. Players' heterogeneity and individuals' choice randomness (the two main components that explain why games outcomes are probabilistic) are somewhat confounded in our formulation, particularly in the parameter of the decision rules, and cannot be analyzed and estimated separately. The major obstacle to confronting this difficulty is getting data sets that are large enough, in the sense of providing a sufficient number of choices per individual.

We focused on the three types of bargaining behavior that the investigations in our sample also emphasized: average opening offers, rejection rates for opening offers and rates of disadvantageous counteroffers. Going further into the bargaining game is difficult, in part, because the amount of data tails off dramatically after the first round of play, owing to first round settlements. That said, there is reason to think that our model is unlikely to have as much success with behavior further into the game: As Bolton (1991) noted, counteroffers in sequential bargaining games often appear to be fashioned responses to previous offers; meaning history matters in the latter rounds in a manner that a model as simple as ours cannot hope to capture.

Finally, it remains to be seen whether fairness estimates are stable across classes of games, beyond bargaining games. Investigating this issue requires a good deal more work, and grappling with some of the issues we have mentioned here. We hope to have something to say on the matter in the near future.

7 REFERENCES

Abbink, Klaus, Bolton, Karim Sadrieh & Fang-Fang Tang (2001), Learning versus Punishment in Ultimatum Bargaining, *Games and Economic Behavior*, 37, 1-25.

Andreoni, James and John H. Miller (2002), Giving According to GARP: An Experimental Test of the Rationality of Altruism, *Econometrica*, 70, 737–54.

Bereby-Meyer, Yoella, and Muriel Niederle (forthcoming), "Fairness in Bargaining," *Journal of Economic Behavior and Organization.*

Binmore, Kenneth, John McCarthy, Giovanni Ponti, Larry Samuelson and Avner Shaked (2002), "A Backward Induction Experiment," *Journal of Economic Theory*, 104, 48-88.

Binmore, Kenneth, Peter Morgan, Avner Shaked and John Sutton (1991), "Do People Exploit Their Bargaining Power? An Experimental Study," *Games and Economic Behavior*, 3, 295-322.

Binmore, Kenneth, Avner Shaked, and John Sutton (1988), "A Further Test of Noncooperative Bargaining Theory: Reply, *American Economic Review*, 75, 1178-80,

Binmore, Kenneth, Avner Shaked, and John Sutton (1985), "Testing Noncooperative Bargaining Theory: A Preliminary Study," *American Economic Review*, 75 (December), 1178-80.

Bolton, Gary E. (1991), "A Comparative Model of Bargaining: Theory and Evidence," *American Economic Review*, 81 (5), 1096-136.

Bolton, Gary E. and Axel Ockenfels (forthcoming), "A Stress Test of Measures of Fairness in Social Utility", *Economic Theory*, http://lems.smeal.psu.edu/lema.

Bolton, Gary E. and Axel Ockenfels (2000), "ERC: A Theory of Equity, Reciprocity, and Competition," *American Economic Review*, 90 (1), 166-93.

Bolton, Gary E. and Axel Ockenfels (1998), "Strategy and Equity: An ERC-Analysis of the Güth-van Damme Game," *Journal of Mathematical Psychology*, 42 (2/3), 215-26.

Carpenter, Jeffrey P. (2003), "Is Fairness Used Instrumentally? Evidence from Sequential Bargaining," *Journal of Economic Psychology*, 24, 467-89.

Charness, Gary and Matthew Rabin (2002). "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics*, 117, 817-869.

Cooper, David J., Nick Feltovich, Alvin E. Roth and Rami Zwick (2003), "Relative versus Absolute Speed of Adjustment in Strategic Environments: Responder Behavior in Ultimatum Games," *Experimental Economics*, 6, 181–207.

Costa-Gomez, M. and K. G. Zauner (2001), "Ultimatum Bargaining Behavior in Israel, Japan, Slovenia, and the United States: A Social Utility Analysis," *Games and Economic Behavior,* 34, 238–70.

Davison, A.C. and D.V. Hinkley (1997), Bootstrap Methods and Their Application. Cambridge Series in Statistical and Probabilistic Mathematics: Cambridge University Press.

Eliason, Scott R. (1993), *Maximum Likelihood Estimation: Logic and Practice*, Sage University Paper on Quantitative Applications in the Social Sciences. Thousand Oaks, CA: Sage.

Fehr, Ernst, Georg Kirchsteiger, and Arno Riedl (1993), "Does Fairness Prevent Market Clearing? An Experimental Investigation," *Quarterly Journal of Economics*, 108 (2), 437-59.

Fehr, Ernst, and Klaus Schmidt (1999). "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics*, 114, 817-868.

Forsythe, Robert, Joel Horowitz, N. E. Savin, and Martin Sefton, (1994), "Fairness in Simple Bargaining Experiments," *Games and Economic Behavior*, *6*(3), 347-69.

Forsythe, Robert, John Kennan and Barry Sopher (1991), "An Experimental Analysis of Bargaining and Strikes with One Sided information," *American Economic Review*, 81, 253-78.

Goeree, Jacob K. and Charles A. Holt (2003), "Asymmetric Inequality Aversion and Noisy Behavior in Alternating-Offer Bargaining Games," http://www.people.virginia.edu/~cah2k.

Goeree, Jacob K., Charles A. Holt, and Susan K. Laury, "Private Costs and Public Benefits: Unraveling the Effects of Altruism and Noisy Behavior," http://www.people.virginia.edu/~cah2k/pubexp.pdf.

Grosskopf, Brit (1999), "Competition, Aspiration and Learning in the Ultimatum Game: An Experimental Investigation," in 1999 European Economics Association Meetings. Universitat Pompeu Fabra.

Güth, Werner , Carsten Schmidt, and Matthias Sutter (2002), "Bargaining Outside The Lab – A Newspaper Experiment Of A Three Person-Ultimatum Game,".

Güth, Werner, R. Schmittberger, and B. Schwarze (1982), "An Experimental Analysis of Ultimatum Bargaining," *Journal of Economic Behavior and Organization*, 3, 367-88.

Güth, Werner and Reinhard Tietz (1988), "Ultimatum Bargaining for a Shrinking Cake – An Experimental Analysis," in R. Tietz, W. Albers, R. Selten, eds., *Bounded Rational Behavior in Experimental Games and Markets*, Springer: Berlin.

Güth, Werner and Eric van Damme (1998), "Information, Strategic Behavior and Fairness in Ultimatum Bargaining, An Experimental Study," *Journal of Mathematical Psychology*, 42 (2/3), 227-47.

Haile, Philip A., Ali Hortaçsu and Grigory Kosenok (2003), "On the Empirical Content of Quantal Response Equilibrium," working paper, http://www.econ.yale.edu/~pah29/qre.pdf.

Harrison, Glenn W. and Kevin A. McCabe (1992), "Testing Noncooperative Bargaining Theory in Experiments," in R. Mark Isaac (ed.), *Research in Experimental Economics, Vol. 5*, JAI Press: Greenwich, Conn., 137-69.

Johnson, Eric J., Colin Camerer, and Sankar Sen (2002), "Detecting Failures of Backward Induction: Monitoring Information Search in Sequential Bargaining," *Journal of Economic Theory*, 104, 1-31.

Kagel, John, Chung Kim, and Donald Moser (1996), "Fairness in Ultimatum Games with Asymmetric Information and Asymmetric Payoffs." *Games and Economic Behavior*, *13*(1), 100-10.

Kagel, John and Katharine Wolfe (forthcoming), "Tests of Difference Aversion Models in Explaining Anomalies in Simple Bargaining Games," *Experimental Economics.*

Louviere, Jordan, Deborah Street, Richard Carson, Andrew Ainslie, J.R. Deshazo, Trudy Cameron, David Hensher, Robert Kohn, and Tony Marley (2002), "Dissecting the Random Component of Utility,: *Marketing Letters*, 13, 177-93.

Luce, Duncan R. (1959), *Individual Choice Behaviour*, New York: Wesley.

Manly, Bryan F. J. (1997), *Randomization, Bootstrap and Monte Carlo Methods in Biology*, Second Edition: CRC Press.

McKelvey, Richard D. and Thomas R. Palfrey, "Quantal Response Equilibria for Normal-form Games," *Games and Economic Behavior*, 7, 6-38.

Mitzkewitz, Michael and Nagel, Rosemarie (1993), "Envy, Greed and Anticipation in Ultimatum Games with Incomplete Information," *International Journal of Game Theory*, *22*(2), 171-98.

Neelin, Janet, Hugo Sonnenschein, and Matthew Spiegel (1988), "A Further Test of Noncooperative Bargaining Theory," *American Economic Review*, 78 (September), 824-36.

Ochs, Jack and Alvin E. Roth (1989), "An Experimental Study of Sequential Bargaining," *American Economic Review*, 79 (June), 355-84.

Rapoport, Amnon, Eythan Weg and Dan S. Felsenthal (1990), "Effects of Fixed Costs in Two-Person Sequential Bargainig," *Theory and Decision*, 28, 47-72.

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

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 (Special Issue: Nobel Symposium), 164-212.

Roth, Alvin E., Vesna Prasnikar, Masahiro Okuno-Fujiwara, and Shmuel Zamir (1991), "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo," *American Economic Review*, 81 (5), 1068-95.

Sarin, Rajiv and Farshid Vahid (1999), "Payoff Assessments Without Probabilities: A Simple Dynamic Model of Choice," *Games and Economic Behavior*, 28, 294-309.

Sarin, Rajiv and Farshid Vahid (2001), "Predicting How People Play Games: A Simple Dynamic Model of Choice," *Games and Economic Behavior*, 34 (1), 104-22.

Slonim, Robert and Roth, Alvin E. (1998), "Learning in High Stakes Ultimatum Games: An Experiment in the Slovak Republic." *Econometrica*, *66*(3), pp. 569-596.

Zwick, Rami, Amnon Rapoport and John C. Howard (1992), "Two Person Sequential Bargaining Behavior with Exogenous Breakdown," *Theory and Decision*, 32, 241-68.

8 APPENDIX

Experiment	Initials	#	Pie size	Rounds	Discount factors	Times played	Offer Obs.	Offer Model	Rejection Obs.	Rejection Model	Disadv. counter. Obs.	Disadv. counter. Model
Roth, Prasnikar, Okuno-Fujiwara and Zamir (1991)	RPOZ	1	$10 or $30	1	n/a	10	.407	.406	.264	.308	n/a	
Binmore, Shaked and Sutton (1985)	BSS	1	100 pence	2	(.25, .25)	1	.416	.416	.148	.497	.750	.740
Güth and Tietz (1988)	GT	1	5 to 35 DM	2	(.10, .10)	1	.281	.370	.190	.381	.750	.914
		2	5 to 35 DM	2	(.90, .90)	1	.427	.357	.619	.551	.000	.130
Neelin, Sonnenschein and Spiegel (1988)	NSS	1	$5	2	(.25, .25)	1	.274	.346	.225	.506	.556	.682
		2	$5	3	(.50, .50)	1	.472	.360	.050	.524	.500	.385
		3	$5	5	(.34, .34)	1	.342	.354	.125	.514	.400	.577
		4	$15	5	(.34, .34)	4	.359	.425	.156	.281	.857	.902
Ochs and Roth (1989)	OR	1	$30	2	(.40, .40)	10	.413	.426	.100	.130	.600	.997
		2	$30	2	(.60, .40)	10	.487	.434	.150	.167	1.000	.997
		3	$30	2	(.60, .60)	10	.473	.463	.188	.202	.733	.909
		4	$30	2	(.40, .60)	10	.457	.446	.200	.170	.550	.984
		5	$30	3	(.40, .40)	10	.433	.422	.120	.137	1.000	.994
		6	$30	3	(.60, .40)	10	.447	.424	.140	.164	.857	.995
		7	$30	3	(.60, .60)	10	.453	.440	.144	.226	.462	.924
		8	$30	3	(.40, .60)	10	.467	.437	.289	.180	.885	.968
Bolton (1991)	B	1	$12	2	(.67, .33)	8	.400	.412	.188	.328	.833	.860
		2	$12	2	(.33, .67)	7	.482	.426	.184	.335	.200	.584
		3	$12	Trunc.	(.67, .33)	8	.407	.469	.391	.370	.960	.906
		4	$12	Trunc.	(.33, .67)	8	.653	.572	.266	.535	.000	.376
Güth and van Damme (1998)	GvD	y	DG 24	3-person	n/a	6	.276 .065	.294 .070	.097	.281	n/a	a
Pearson's R correlation	(All observations)							.898		.331		.820
	(Excluding RPOZ and GvD)							.771		.331		.820
	(Excluding RPOZ, GvD and small pie games)							.847		.688		.856

(a) Top number refers to mean offer to the responder, and bottom mean offer to the dummy.

Table 7 – Opening offer and rejection behavior by treatment: observations and model predictions.