MANAGEMENT SCIENCE
Vol. 54, No. 10, October 2008, pp. 1774–1791
ISSN 0025-1909 | EISSN 1526-5501 | 08 | 5410 | 1774

informs®

DOI 10.1287/mnsc.1080.0887

Estimating the Influence of Fairness on Bargaining Behavior

Arnaud De Bruyn
ESSEC Business School, 95000 Cergy, France, debruyn@essec.fr

Gary E. Bolton
Smeal College of Business, Pennsylvania State University, University Park, Pennsylvania 16802, gbolton@psu.edu

The strength of bargainers' preferences for fair settlements has important implications for predicting negotiation outcomes and guiding bargaining strategy. Existing literature reports a few calibration exercises for social utility models, but the predictive accuracy of these models for out-of-sample forecasting remains unknown. Therefore, we investigate whether fairness considerations are stable enough across bargaining situations to be quantified and used to forecast bargaining behavior accurately. We develop a model that embeds a preference for fair treatment in a quantal response framework to account for noise and experience. In addition, we estimate preference for fairness (willingness to pay) using the simplest, one-round version of sequential bargaining games and then employ it to perform out-of-sample forecasts of multiple-round games of various lengths, discount factors, pie sizes, and levels of bargainer experience. Except in circumstances in which the bargaining pie is very small, the fitted model has significant and substantial out-of-sample explanatory power. The stability we find implies that the model and techniques might ultimately be extended to estimates of the influence of fairness on field negotiations, as well as across subpopulations.

Key words: games; group decisions; bargaining; utility-preference; estimation
History: Accepted by George Wu, decision anlysis; received April 13, 2006. This paper was with the authors 1 year and 3 months for 2 revisions. Published online in *Articles in Advance* July 31, 2008.

1. Introduction

Although bargainers commonly demand fair settlements, estimating their preferences for fairness has proven difficult, largely because demands for fairness are easily confounded with other strategic objectives of the negotiator. Thus, it can be difficult to discern whether fairness preferences shape the negotiation or whether the negotiation shapes fairness preferences.

We develop a strategic model of bargainer behavior that includes both fairness and monetary preferences and investigate the model's out-of-sample forecasting properties. We find that the estimates of fairness preferences obtained in one bargaining situation are indicative of bargainer behavior in other situations, where strategic parameters such as time cost, stake, and length of negotiations differ. By demonstrating both a method for estimating fairness preferences and the stability of the estimates obtained, with further research, it may be practical to quantify and embed fairness preferences into sales or negotiation decision support systems.[1]

Several studies include calibration exercises for social utility models (e.g., Goeree and Holt 2000, Costa-Gomes and Zauner 2001, Fehr and Schmidt 1999, Bolton and Ockenfels 2000, Charness and Rabin 2002). Our study differs from these in that we fit an ultimatum game in an attempt to extend the model to out-of-sample estimates for other bargaining games, which in turn enables us to test the stability of willingness to pay for fair treatment across different bargaining environments. Our research goals thus are threefold:

1. Demonstrate that preference models that incorporate fairness considerations explain bargaining behavior better than traditional preference models;
2. Explore whether a social utility model calibrated on one bargaining data set can make reliable, out-of-sample forecasts in bargaining environments with different characteristics; and
3. Compare the fit obtained from alternative specifications of the preference function.

The remainder of this paper is organized as follows: In §2, we describe the structure of bargaining games, as well as the data we use for both in-sample fit and out-of-sample exploration. Then, in §3, we lay out the *equity–reciprocity–competition* (ERC) model (Bolton and Ockenfels 2000) that we employ to fit the data. Next, in §4, we estimate the ERC model with the Roth et al. (1991) ultimatum game data and, in §5, use that model

[1] Our approach is in line with the research agenda for fairness proposed by Narasimhan et al. (2005).

to obtain out-of-sample estimates for multiple-round bargaining games. The model provides satisfactory fit along several dimensions, including first-offer behavior, rejection behavior, disadvantageous counteroffers, experience trends, and empirical regularities. In §6, we check the robustness of our proposed model's performance with respect to (1) alternative specifications of the fairness component of the model, (2) alternative fits using data from two- and three-round games, and (3) a three-person bargaining situation that stress-tests the model according to the presumed fair standard and an extreme settlement outcome, unobserved in two-person sequential bargaining games. In §7, we conclude and discuss the implications of our research.

2. Sequential Bargaining Game Data

2.1. Sequential Bargaining Games

Sequential bargaining involves an offer–counteroffer format, common to many negotiations. In a sequential bargaining game, a buyer α and a seller β seek to split a pie of c units. Conceptually, this pie represents the difference between the buyer's and the seller's reservation price (e.g., Samuelson 1980); any offer between these two limits provides a positive surplus to both bargainers. The negotiation consists of n rounds. In odd rounds, α chooses x of the pie to offer β, keeping $c - x$ for self. If β accepts, the pie is divided accordingly, but if β rejects the offer, the pie shrinks according to the discount factors $(\delta_\alpha, \delta_\beta)$, which represent the costs of a rejection. If $\delta_\alpha \neq \delta_\beta$, the pie is shrinking at a faster pace for one player than for the other, meaning that the consequences of rejecting an offer vary across players. In even rounds, the roles reverse, so β makes the offer. When they cannot reach an agreement after n rounds, both bargainers receive nothing.

An *ultimatum game* refers to a single-round version of this game, whereas a *truncation game* consists of two rounds, in which α must accept any second-round offer. Finally, the *three-person ultimatum game* is a variant of the one-round version, such that one player offers a split of the pie to three participants, the second player either accepts or rejects that offer, and a third player (the dummy) has nothing to say.

Selfish game theory analysis assumes that each bargainer prefers more money to less, with no fairness considerations. The selfish (subgame perfect) equilibrium is constructed by backward induction. To illustrate, suppose the bargaining pie equals \$10. Consider first the ultimatum game. Because β prefers more money, α should offer the smallest unit possible, say \$1, which β should accept. Building on this, consider a two-round negotiation, and observe that the second round is an ultimatum game in which β is the proposer. If the game goes to the second round, in equilibrium, β offers \$1. Knowing this offer will come, α needs to offer an amount in the first round just slightly larger than β can expect in round 2. Consequently, the game should end in round 1 with β receiving $\delta_\beta 10$ and α receiving $(1 - \delta_\beta)10$. A similar inductive technique can construct the equilibrium for any n-round game.

Fairness enters the picture because of the systematic ways in which actual play deviates from the selfish equilibrium. For example, first-round offers tend to deviate toward greater equity. In addition, more inequitable (but closer to equilibrium) offers often get rejected. Rejections in multiple-round games often prompt disadvantageous counteroffers, that is, the rejection of an offer, followed by a counteroffer that gives the proposer less in absolute terms than the original offer would have. As we discuss below, the extent of these deviations varies greatly depending on the version of the game, the number of rounds, the discount factors, and so forth.

2.2. Sequential Bargaining Data

The data we use for our investigation encompass all sequential bargaining game studies included in the "Bargaining Experiments" chapter of the *Handbook of Experimental Economics* (Roth 1995). Altogether, there are 2,726 observations from 1,037 participants, collected in 21 different experimental conditions by seven distinct research teams in seven countries on three continents. These influential studies have inspired a host of research in a variety of directions. Two additional attributes make this collection of data particularly attractive for our purposes. First, sequential bargaining not only represents a major paradigm for the study of bargaining in management literature[2] but also offers a simple and unified theoretical structure, which assists our efforts to fit all the data to a single model. Second, the diversity of game parameterizations encompassed by these data sets provides a rich test bed for challenging a model of fairness preferences. The role of fairness was a central point of contention among these studies, and each study can be considered a robustness test of the claims made by the preceding studies in the set.

Table 1 illustrates the variability within the *Handbook* sequential bargaining studies along a number of dimensions, starting with game parameterization. Monetary pie sizes, indicative of the incentives offered for subject participation, differ substantially across experiments and, because of the international nature of the data set, were offered in a variety of currencies. In addition, the number of rounds per

[2] Among many examples, Rapoport et al. (1995) use such mechanisms to analyze the influence of one-sided, incomplete information in bargaining situations, and Gächter and Riedl (2005) study moral property rights and the influence of infeasible claims on negotiations.

game varies from one to five. Discount factors used also vary across the feasible range, from 0.1 to 0.9. (The three-person ultimatum game studied by Güth and van Damme 1998, also listed in Table 1, will be used in a robustness test of the model in §6.)

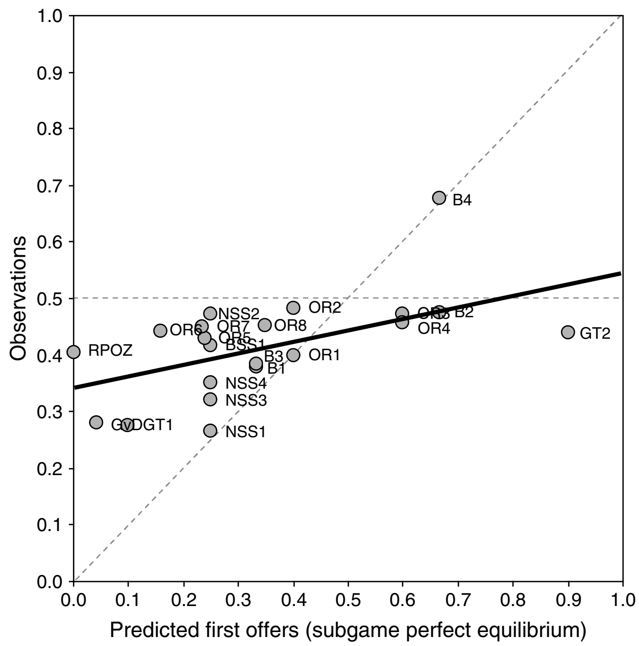
Table 1 also illustrates the diversity in experimental designs. First, the number of treatments that each experiment performed varies greatly; Ochs and Roth (1989) investigate the most in this regard. Second, a correspondingly wide variation appears in the number of subjects sampled (implicit in the table). For example, the Roth et al. (1991) study took place across four countries. Multiple-round studies take place in England, Germany, and the United States; Güth and van Damme (1998) conduct their study in The Netherlands. Third, the amount of experience with the game (i.e., times played) varies from 1 to 10 games. In all cases though, bargainers play with a given partner only once. Table 1 also displays the selfish equilibrium first offers/settlements. They vary from nearly 0% to 90% of the pie allocated to β. Selfish equilibrium also predicts that no rejection occurs in response to a positive offer.

Taken together, these studies provide a challenging test for an out-of-sample forecasting exercise, especially in terms of the variability in the data. Table 1 lists three measures of bargaining behavior reported in these studies and forecasted in this paper. Average first offers in sequential bargaining games range from 27% to 65%; rejection rates range from 5% to 62%; and disadvantageous counteroffers span 0%–100% (i.e., an offer is rejected and eventually followed by a counteroffer that leads to a lower payoff in absolute terms). This variability occurs even across games that share the same structure. For example, Binmore et al. (1985) and Neelin et al. (1988) both run two-round bargaining games with identical (0.25, 0.25) discount factors, yet the former reports a much higher average first offer (0.416 versus 0.274)—a finding that our study replicates and that we explain by the difference in pie sizes.

In addition, selfish equilibrium offers little help in explaining the differences across studies. In Figure 1, we depict a comparison between the observed average opening offers in these studies and the selfish equilibrium prediction (labels in the figure mirror labels in Table 1). The observed average opening offers for multiple-round games vary noticeably (27%–67%), though not by as much as predicted (10%–90%). We also note a strong first-mover advantage (cf. one Bolton truncation game—which we replicate below). Regressing the observed average opening offers on those predicted by selfish equilibrium yields the following results (two-sided p-values in parentheses):

$$\text{Obs.} = \underset{(0.006)}{0.234}\ \ \text{Pred.} + \underset{(0.000)}{0.334}. \tag{1}$$

Figure 1 **Predicted First Offers Across Multiple-Round Games, Observations vs. Subgame Perfect Model**



Notes. The solid line in the graph is the regression line. See Table 1 for treatment label interpretations.

That is, selfish equilibrium explains only 24.0% of the variance in observed opening offers. Moreover, we easily reject the hypothesis that the coefficient of PREDICTED equals 1 (one-tailed $p < 0.001$). A bias toward higher-than-predicted offers is clearly evident in the intercept term of 0.334.

Several of the studies we investigate find trends in the opening offers, across repeated plays of the game (also see Roth and Erev 1995). In some cases the offers tend to move toward the selfish equilibrium whereas in others they tend to move away. Our analysis indicates that, in most cases, the trend is toward the equilibrium offers implied by fairness preferences, specifically, the ERC equilibrium that we describe next.

3. The Model

Models that incorporate a fixed preference for fairness generally fit with ordinal regularities in the data, though to date no model claims a satisfactory quantitative fit. We adopt Bolton and Ockenfels's (2000) ERC specification[3] as the foundation of our model; ERC hypothesizes that bargainers have preferences over relative payoffs (fairness) as well as absolute payoffs (pecuniary). Bargainers' utility increases as the absolute payoff increases, but it diminishes as their relative payoff deviates from the fair standard. In line with Zwick and Chen (1999), we assume that

[3] In §6, we explore the Fehr–Schmidt preference specification as a comparison.

Table 1 Experimental Designs and Observations—Average First Offers, Rejection Rates, and Disadvantageous Counteroffers—for All Bargaining Studies in the Sample

Experiment	Initials	No.	Pie size	Rounds	Discount factors	Subjects	Times played	Obs.	First offer	GT predict.	Rejection rate	Disadvantag. counteroffers
Roth et al. (1991)[a, b]	RPOZ	1	$10	1	n/a	250	10	1,250	0.395	0.001	0.272 (340/1,250)	n/a
Binmore et al. (1985)[c]	BSS	1	100 pence	2	(0.25, 0.25)	163	1	81	0.416	0.250	0.148 (12/81)	0.750 (9/12)
Güth and Tietz (1988)[d]	GT	1	5 to 35 DM	2	(0.10, 0.10)	42	1	21	0.276	0.100	0.190 (4/21)	0.750 (3/4)
		2	5 to 35 DM	2	(0.90, 0.90)	42	1	21	0.440	0.900	0.619 (13/21)	0.000 (0/13)
Neelin et al. (1988)[e]	NSS	1	$5	2	(0.25, 0.25)	80	1	40	0.265	0.250	0.225 (9/40)	0.556 (5/9)
		2	$5	3	(0.50, 0.50)	80	1	40	0.472	0.250	0.050 (2/40)	0.500 (1/2)
		3	$5	5	(0.34, 0.34)	80	1	40	0.320	0.250	0.125 (5/40)	0.400 (2/5)
		4	$15	5	(0.34, 0.34)	30	4	60	0.348	0.250	0.167 (7/60)	0.857 (6/7)
Ochs and Roth (1989)	OR	1	$30	2	(0.40, 0.40)	20	10	100	0.399	0.400	0.100 (10/100)	0.600 (5/10)
		2	$30	2	(0.60, 0.40)	20	10	100	0.482	0.400	0.150 (15/100)	1.000 (15/15)
		3	$30	2	(0.60, 0.60)	16	10	80	0.471	0.600	0.187 (15/80)	0.733 (11/15)
		4	$30	2	(0.40, 0.60)	20	10	100	0.458	0.600	0.200 (20/100)	0.550 (11/20)
		5	$30	3	(0.40, 0.40)	20	10	100	0.429	0.240	0.120 (12/100)	1.000 (12/12)
		6	$30	3	(0.60, 0.40)	20	10	100	0.443	0.160	0.140 (14/100)	0.857 (12/14)
		7	$30	3	(0.60, 0.60)	18	10	90	0.449	0.235	0.144 (13/90)	0.462 (6/13)
		8	$30	3	(0.40, 0.60)	18	10	90	0.453	0.350	0.289 (26/90)	0.885 (23/16)
Bolton (1991)[f]	B	1	$12	2	(0.67, 0.33)	16	8	64	0.378	0.333	0.188 (12/64)	0.833 (10/12)
		2	$12	2	(0.33, 0.67)	14	7	49	0.476	0.666	0.204 (9/49)	0.200 (2/9)
		3	$12	Trunc.	(0.67, 0.33)	16	8	64	0.384	0.333	0.391 (25/64)	0.960 (24/25)
		4	$12	Trunc.	(0.33, 0.67)	16	8	64	0.678	0.666	0.266 (17/64)	0.000 (0/17)
Güth and van Damme (1998)[b, g]	GvD	y	DG 24	Three-person	n/a	36	6	72	0.276 0.065	0.042 0.042	0.097 (7/72)	n/a

[a]Numbers reported are aggregations of four-treatment run, respectively, in Israel, Japan, Slovenia, and the United States. Payoffs are in local currency; size of pie outside of the United States 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 the bottom number refers to mean offer to the dummy. Minimum offer allowed: 5 tokens to each player (out of 120).

"fairness" has a price and aim to estimate it. Thus, bargainers reject extreme offers if the expected consequence is less than their willingness to pay, which depends on the parameters of the game. We investigate a simple version of the model that estimates the average trade-off bargainers face between fairness and material gain.

To account for experience effects and noise in behavior, we insert ERC preferences into a quantal response equilibrium framework (McKelvey and Palfrey 1995). Most choice theories proposed in psychology assume that choice behavior is probabilistic (e.g., Luce 1959). Quantal response permits "mistakes" with respect to the optimal decision. The working hypothesis behind our model is that noise diminishes with experience, moving behavior closer to the static ERC equilibrium. We emphasize that we do not consider the quantal response framework to be a model of learning; rather, we think of it as a relatively straightforward technique for accounting for noise and experience. Finally, our model supposes that "certainty of choice" is greater with larger stakes.

3.1. The ERC Preference Function

The utility function we use is similar to that suggested by Bolton and Ockenfels (2000) but is restricted to the kind of asymmetry suggested by Bolton's (1991) comparative bargaining model. Specifically,

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

where c is the size of the pie, σ is the proportion of the pie the player receives, and b measures the relative importance of any deviation from an equitable allocation. The absolute and relative payoffs are additively separable, such that relative payoff entails an asymmetric loss function, with no loss (0) when the player obtains half or more of the bargaining pie (i.e., the two-person game's fair standard is an equal division). The marginal utility of the relative component grows greater as the player's share drops farther below half. The function contains one fitted parameter, b, whose value likely varies across different players, though we interpret our estimate as the population average. Finally, $U(\sigma)$ ranges between $-\infty$ (if $b=\infty$ and $\sigma < 1/2$) and c.

The asymmetry of the utility function refers to bargaining game data, not the character of the utility function per se. Some researchers refer to bargaining games as games of negative reciprocity because the bargainers' tendency to reject unfair offers dominates the tendency to make fair offers, in the sense that the former becomes the major influence on proposers' offers (e.g., Forsythe et al. 1994). The concept of negative reciprocity also links to the idea of envy discussed in management literature (Boiney 1995). As we show below, it is difficult to distinguish the completely asymmetric formulation of Equation (2) statistically from symmetric formulations, and the asymmetric formulation provides better out-of-sample forecasts than a strictly symmetric formulation.

In our formulation, the size of the bargaining pie c does not affect the weights that a bargainer gives to relative versus absolute payoffs. That is, players remain driven in the same proportion by absolute and relative payoffs regardless of pie sizes, though a player prefers a given share from a larger pie more than from a smaller one. Nevertheless, in this model, pie size influences decisions.

3.2. Decision-Making Framework: Quantal Response

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

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

$$P_\beta(\sigma_i) = \frac{e^{\tau_\beta \cdot U(\sigma_i)}}{e^{\tau_\beta \cdot U(\varnothing)} + e^{\tau_\beta \cdot U(\sigma_i)}} = \frac{e^{\tau_\beta \cdot U(\sigma_i)}}{1 + e^{\tau_\beta \cdot U(\sigma_i)}}, \qquad (3)$$

where $U(\varnothing)$ and $U(\sigma_i)$ are the utilities of rejecting or accepting the offer (indiced i), respectively, calculated from Equation (2). A rejection shrinks the size of the pie to $c=0$, so $U(\varnothing)=0$ for all σ_i.

We refer to $\tau_\beta > 0$ as the coefficient of certitude, such that τ_β indicates the players' choice consistency. As $\tau_\beta \to \infty$, if $U(\sigma_i) > U(\varnothing)$, then $P_\beta(\sigma_i) \to 1$; however, if $U(\sigma_i) < U(\varnothing)$, then $P_\beta(\sigma_i) \to 0$. That is, the larger τ_β, the greater is the probability that the responder follows a strategy that produces the highest utility. At the other extreme, if $\tau_\beta = 0$, there is an equal chance the responder will accept or reject the offer, independent of the actual value of σ_i, indicating that offers get accepted or rejected arbitrarily. Various experiments show that players can be inconsistent over time, accepting an offer in one game but refusing a 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 creates some uncertainty about which strategy the same player will follow over time.

3.2.2. 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 model, the proposer's decision to offer σ_i (and keep $1-\sigma_i$ for self) follows a logit distribution:

$$P_\alpha(\sigma_i) = \frac{e^{\tau_\alpha \cdot \mathrm{E}(U(1-\sigma_i))}}{\sum_{j=1}^{I} e^{\tau_\alpha \cdot \mathrm{E}(U(1-\sigma_j))}} = \frac{e^{\tau_\alpha \cdot P_\beta(\sigma_i) . U(1-\sigma_i)}}{\sum_{j=1}^{I} e^{\tau_\alpha \cdot P_\beta(\sigma_j) . U(1-\sigma_j)}}, \qquad (4)$$

where $\mathrm{E}(U(1-\sigma_i))$ 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 with which responders will accept any particular offer, so the expected utility of an offer σ_i equals $P_\beta(\sigma_i) \cdot U(1-\sigma_i)$.

It then follows that $\sum_i P_\alpha(\sigma_i) = 1$, and offers with the highest expected utilities should be chosen more often. Again, as $\tau_\alpha \to \infty$, proposers systematically make an offer σ_i that procures the highest expected utility. However, if $\tau_\alpha = 0$, $P_\alpha(\sigma_i) = P_\alpha(\sigma_j)$ for all i and j.

3.2.3. Modeling the Coefficient of Certitude. We model the coefficient of certitude as

$$\tau_\alpha = \tau'_\alpha(1+\tau_1 g) \qquad (5)$$

and

$$\tau_\beta = \tau'_\beta(1+\tau_1 g), \qquad (6)$$

where τ'_α and τ'_β are the base decision parameters for players α and β, respectively, τ'_1 is the experience trend, and g is the number of games played. Parameters vary across player roles, because the nature of their decisions differ substantially (e.g., how much to offer versus accept or reject). Because coefficients of certitude tend to increase mechanically with the number of options available, we intuitively expect τ_α (one-of-many decision) to be greater than τ_β (binary decision). We also discount the coefficient of certitude according to the amount of experience the bargainer has, such that the decision parameter, and decision certainty, increases with experience. Moreover, we assume τ_1 to be identical for α an β, and both players' behavior changes at a similar rate. Although this assumption may not hold, it makes the model more parsimonious. In multiple-round games, players can make counteroffers and thereby switch roles during a game; assuming a constant rate of learning across roles simplifies the model tremendously.[4]

[4] We also tested whether the model provides a better fit if we estimate two rates of learning. The differences are not significant at $p < 0.05$.

As Equations (2)–(4) reveal, increasing the size of the pie has the same effect as increasing the coefficient of certitude. Therefore, the model structure clearly conveys that players select those actions with the greatest expected rewards more frequently when the size of the pie is larger; players' decisions should be more consistent and less erratic when the game's stakes are higher.

4. Fitting the Model to Ultimatum Game Data

4.1. Selecting a Data Set to Calibrate the Preference Model

In terms of model calibration, the Roth et al. (1991) ultimatum game experiment offers several features that make it nearly ideal for fitting the model initially and then forecasting the sequential bargaining games out of sample. First, bargainers play repeatedly, which we need to estimate experience effects (which Roth et al. 1991 find). Second, their paper represents one of the largest ultimatum game studies, with some 1,350 observations, that includes many observations per round. We fit our model using maximum likelihood estimation, and desirable asymptotic properties are justified only in large sample situations (Eliason 1993). Third, Roth et al. (1991) gather their data from four countries (Israel, Japan, Slovenia, and the United States) and find modest differences in bargainer behavior across countries. Data for the multiple-round games come from three different countries (Germany, Great Britain, and the United States). Our model does not control for cultural differences, but fitting it with a multi-country data set should avoid bias in the estimates, which we would expect with single-country data.

4.2. Estimating the Parameters of the Model from the Ultimatum Game Data

The model contains four parameters to be estimated: b, the unique parameter of the utility function; τ'_α and τ'_β, the two base decision parameters; and τ_1, the experience trend parameter. The game design determines the other parameters, namely, c and g.

We fit the model to the multi-country bargaining experiment conducted by Roth et al. (1991) using maximum likelihood estimation. Because we assume nothing about the underlying process that generated the data, we estimate the standard deviations using nonparametric bootstrap variance estimation (see Davison and Hinkley 1997). Bootstrap variance estimation also enables us to estimate the standard deviation of the log-likelihood, which we use subsequently to compare the different model specifications. The parameter estimates we obtain are as follows (standard deviations in parentheses):

$$\begin{array}{cccc} b = 6.692 & \tau'_\alpha = 0.690 & \tau'_\beta = 0.280 & \tau_1 = 0.065 \\ (0.740) & (0.073) & (0.027) & (0.016) \end{array}.$$

All parameters have the expected sign and are significant at $p < 0.01$. The model predicts that the average offer will be 39.2%[5] of the pie, with an average rejection rate of 27.5%. The corresponding actual observations are 39.5% and 27.2%. Student t-tests of these two measures at $p < 0.05$ cannot reject the model. The log-likelihood is −2,694 (s.d. 44.8), and the correlations between the observations and predictions are high, with $R_\alpha = 0.898$ and $R_\beta = 0.972$. We depict the fit in Figure 2.

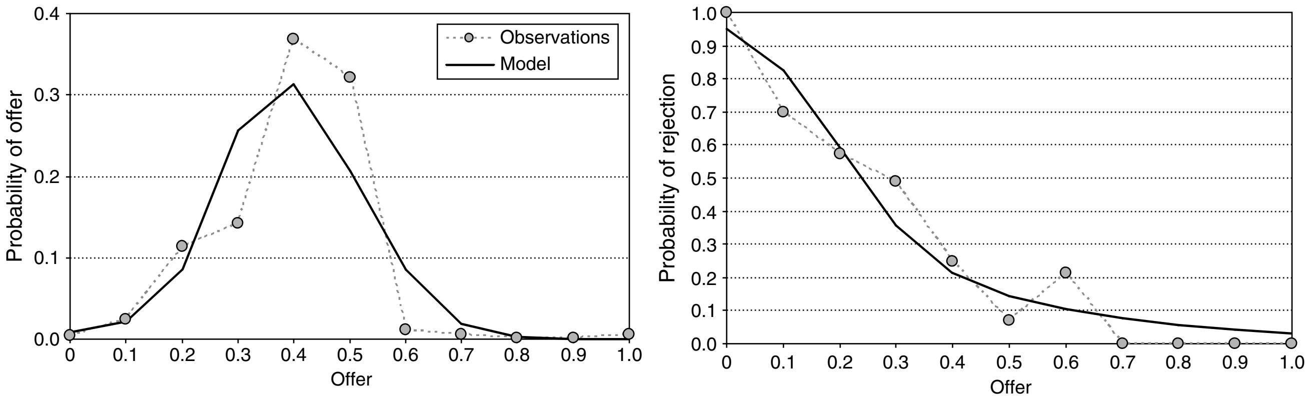
In the next two figures, we display how experience affects both the mean opening offers and rejection rates. Figure 3 reports the first-offer distributions (model versus observations) in the first versus the last rounds of the experiment, after some experience. According to the model, offers tend to tighten around 40% because of the reduction in the amount of error that proposers make in determining the optimal offer, given the population probability of rejection.

Figure 4 reports the average rejection rate during the 10 rounds of the game. Rejection rates reach their lowest levels during the last rounds, a trend effectively captured by the model (which attributes this change to the more generous offers that proposers tend to make as they gain experience).

Recall that the ERC model tested here assumes perfect asymmetry with respect to the relative payoff; specifically, a bargainer cares only about fairness for self. This is not to say that we have evidence against caring about fairness toward others. The log-likelihood of the asymmetric model is −2,694, with a standard deviation of 44.8 (bootstrap variance estimate). The log-likelihood of the model with perfect symmetry is −2,651, with a standard deviation of 32.3. Although the fit statistically improves at the margin ($p < 0.1$), this symmetric specification probably captures noise in the data,[5] so we retain the original, asymmetric model. As we reveal in §6, there are several reasons to prefer it to a symmetric one.

[5] For example, some participants offer the entire pie to the responder, especially during their first play, and the latter reject the offer. Although this kind of behavior might be explained by very strong fairness concerns on both sides (preferring receiving nothing than an unfair share), it is more likely that some participants were confused about what they had to write down (i.e., the share they offered to keep or the share they offered to give) the first time they played the game. The asymmetric specification suggests very low probabilities for these events, so they affect the log-likelihood function significantly. Removing anomalies from the data leads to no differences in the likelihoods.

Figure 2 Probability That the Proposer Makes an Offer of σ_i (Left) and Probability That the Responder Rejects Such an Offer (Right), Observations vs. ERC Model



Source. Roth et al. (1991).

5. Out-of-Sample Fit to Multiple-Round Sequential Bargaining Games

5.1. Procedure

In computing out-of-sample forecasts, the specific value of the coefficient of certitude in any given decision varies with the type of decision made, not with the player who makes it. That is, when a player makes an offer (or counteroffer), he faces a one-option-out-of-many kind of decision and takes a coefficient of certitude of τ'_α; when he responds to an offer (or counteroffer), he faces a binary decision and takes a coefficient of certitude of τ'_β.

We estimate the solutions by backward induction. Specifically, we build a full decision tree in which each node represents a decision ("make an offer of x" or "accept/reject the offer") and each leaf has an associated probability of occurrence. We first compute the utilities and associated choice probabilities for each end leaf, whose actual utilities are provided by the model. To compute the expected utilities and choice probabilities for parent nodes, we use backward induction and continue until we reach the first node of the tree. We average the forecasts of repeated games over an equal number of games (with an experience trend) as in the comparison experiment.

In the appendix, we provide a detailed accounting of the observations and associated predictions. Finally, we examine the opening offers, rejection rates of opening offers, disadvantageous counteroffers, learning trends, and other regularities reported by various authors in previous studies.

5.2. Opening Offers

Figure 5 plots predicted average opening offers against actual observations in each multiple-round study. The figure also plots the associated regression line.

A perfect fit is along the 45° line. We can see that the model fits these data quite well. Specifically, regressing the observations on predictions, we find (two-sided p-values in parentheses):

$$\text{Obs.} = \underset{(0.000)}{1.517} \quad \text{Pred.} - \underset{(0.137)}{0.214}. \tag{7}$$

The estimated coefficient for PREDICTED is 1.517, and we cannot reject a model that restricts this value to 1 (F-test, $p = 0.232$). The small intercept term is not significant at $p < 0.1$ and indicates little in the way of bias in the estimates. We judge the amount of variability explained as good ($R^2 = 0.596$), given the simplicity of the model and the great variety of bargaining parameters in the sample. Furthermore, the results are robust to other estimation approaches (see the online appendix, provided in the e-companion,[6] for more details).

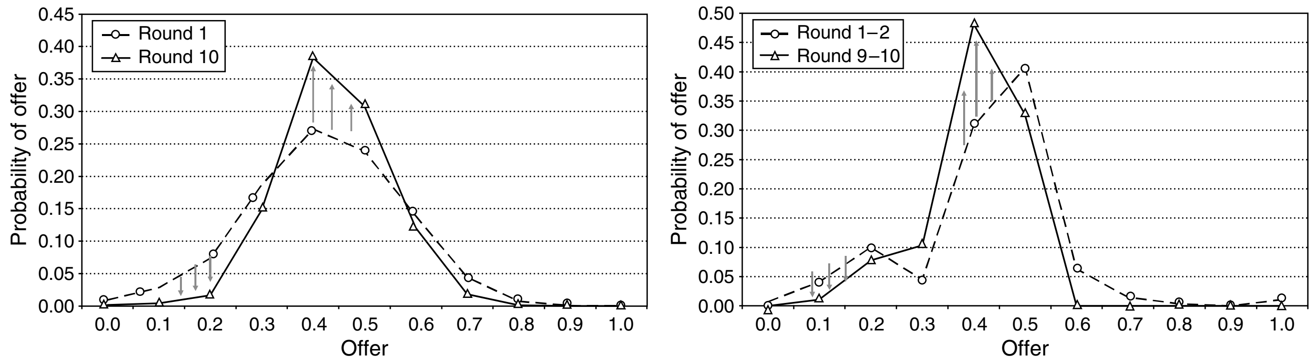
If we drop the extreme observation (truncation game), the regression estimate is ($R^2 = 0.291$; two-sided p-values in parentheses):

$$\text{Obs.} = \underset{(0.021)}{1.069} \quad \text{Pred.} - \underset{(0.852)}{0.033}. \tag{8}$$

The model captures several regularities that were noted by the authors of the sample studies. In Table 1, as Ochs and Roth (1989) note, opening offers tend to be higher than the selfish equilibrium. This regularity can be observed in 14 of the 19 observations and is well captured by the model (predictive accuracy: 93%, $p < 0.01$). Opening offers also tend to be less than half the pie, as both observed and correctly predicted in 18 of the 19 observations (predictive accuracy: 100%, $p < 0.01$). The one exception, also captured

[6] An electronic companion to this paper is available as part of the online version that can be found at http://mansci.journal.informs.org/.

Figure 3 **Probability That the Proposer Makes an Offer of σ_i During the First and Last Periods for the Model (Left) vs. Observations (Right)**



Source. Roth et al. (1991).
Note. Small offers are likely to be rejected, so bargainers eventually make offers that tend to converge around 40% of the pie.

by the model, is a truncation game, in which the offer is considerably higher. Ochs and Roth (1989) also note that the first mover's discount factor affects the outcome (fixing other discount factors), even in a two-round game. As we show in the appendix, for all relevant comparisons, our model predicts that the average opening offers move in the same direction as the data indicate.

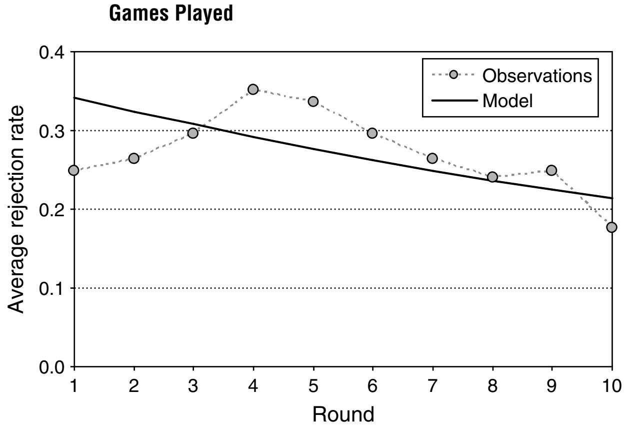
Neelin et al. (1988) observe different behavior in their two-round game than Binmore et al. (1985) observe in theirs, even though the two games use the same discount factors and share the same selfish equilibrium. Our model correctly predicts higher opening offers in the Binmore et al. (1985) games (0.441 observed versus 0.416 predicted) than in the Neelin et al. (1988) games (0.353 observed versus 0.265 predicted) as a result of the differences in the bargaining pie sizes.

5.3. Rejection Rates

In Figure 6, we provide the out-of-sample plots for rejections of opening offers. Again, a perfect fit occurs along the 45° line, though we observe two deviations. First, heteroscedasticity exists, such that variability increases as the predicted rejection increases. Second, for the four games with the smallest bargaining pies, the predicted rejection rates hover around the arbitrary level, or 50%, which is substantially higher than observed. The regression that estimates observations on the basis of predictions, using weighted least squares, is insignificant and accounts for only 15% of the data. However, it remains apparent from Figure 6 that this insignificance is due to the poor small pie predictions.

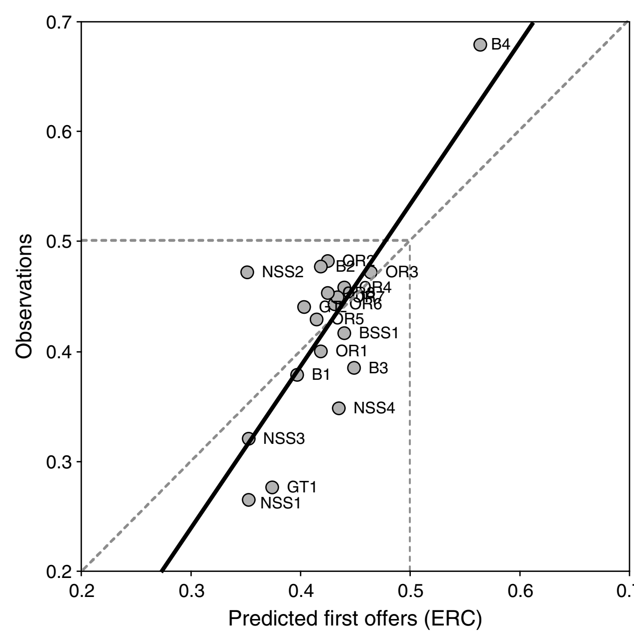
The weighted least squares regression estimate in Figure 6 also adds a dummy variable for the four smallest pie games. It fits the data fairly well, with an

Figure 4 **Average Rejection Rate Diminishes with Number of Games Played**



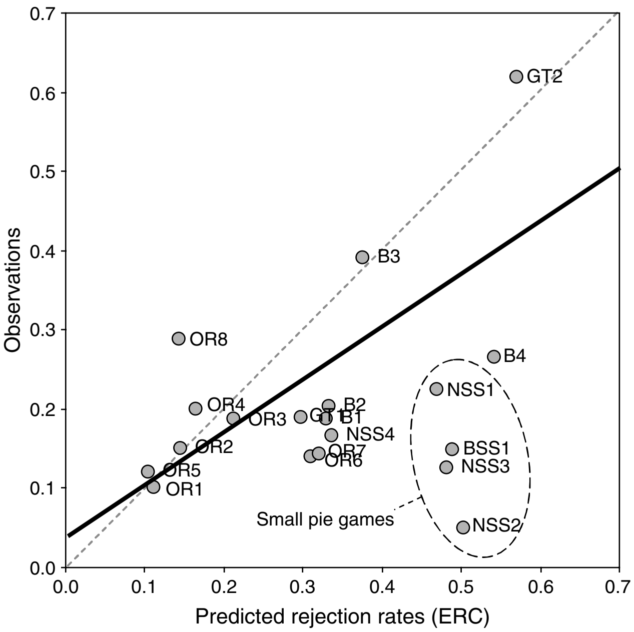
Source. Roth et al. (1991).

Figure 5 **Predicted First Offers Across Multiple-Round Games, Observations vs. Model**



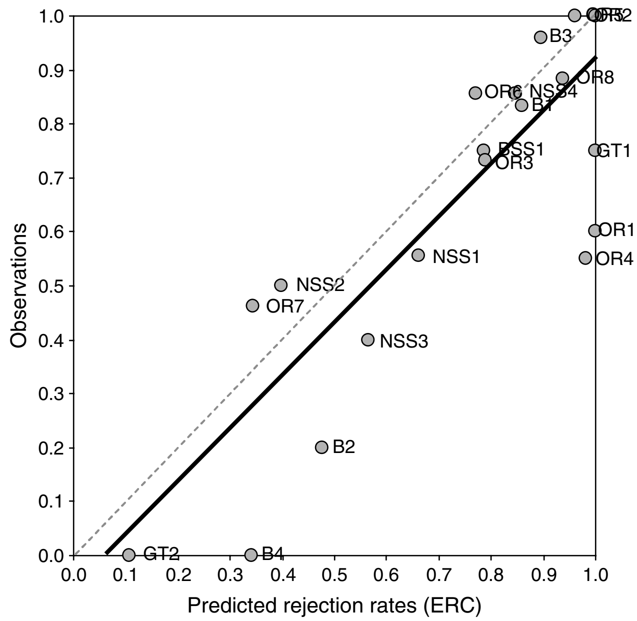
Notes. Solid line is the regression line. See Table 1 for treatment label interpretations.

Figure 6 **Predicted Rejection Rates Across Multiple-Round Games, Observations vs. Model**



Notes. Solid line is the regression line. See Table 1 for treatment label interpretations.

Figure 7 **Predicted Disadvantageous Counteroffers Across Multiple-Round Games, Observations vs. Model**



Notes. Solid line is the regression line. See Table 1 for treatment label interpretations.

R^2 of 0.474 (two-sided p-values in parentheses):

$$\text{Obs.} = \underset{(0.005)}{0.597}\ \text{Pred.} - \underset{(0.007)}{0.206}\ \textit{Small} + \underset{(0.384)}{0.052}, \tag{9}$$

where *Small* is a dummy equal to 1 for small pie games and 0 otherwise. The coefficient for PREDICTED is 0.597. We cannot reject the model that restricts this value to 1 (F-test, $p = 0.183$), and the intercept term is small and not significant. When we drop the small pie games from the sample and rerun the regression, we obtain similar estimates ($R^2 = 0.495$). We might suspect that the smaller the pie, the more likely the model overpredicts, but further regression analysis does not bear out this explanation. Neither pie size nor reciprocal pie size variables reveal any significant correlation with the observed rejection rates, nor does either variable add to the predictive value of the regression. The results are robust to other estimations approaches (see the online appendix).

Güth and Tietz (1988) observe a dramatic rise in the rejection rates when the cost of disagreement decreases; our model generally supports this as a regularity. Ochs and Roth (1989) compare 10 scenarios (fixed number of rounds) in which the cost of disagreement increases (i.e., discount factor decreases and the other factor is fixed, or both decrease). In all cases, the out-of-sample estimates suggest that rejection rates should rise; the observed rejection rates increase in 8 of 10 cases ($p = 0.055$, one-tailed sign test).

5.4. Disadvantageous Counteroffers

We estimate the incidence of disadvantageous counteroffers, when responders reject an initial offer, then eventually make a counteroffer that leads to a lower monetary payoff. Although fewer games continue to the second round (between 2 and 26 per treatment, with an average of 12.4), which means we compute the advantageous/disadvantageous counteroffer proportions on just a few observations, the predictions fit quite well (see Figure 7), with an R^2 of 0.694 (two-sided p-values in parentheses):

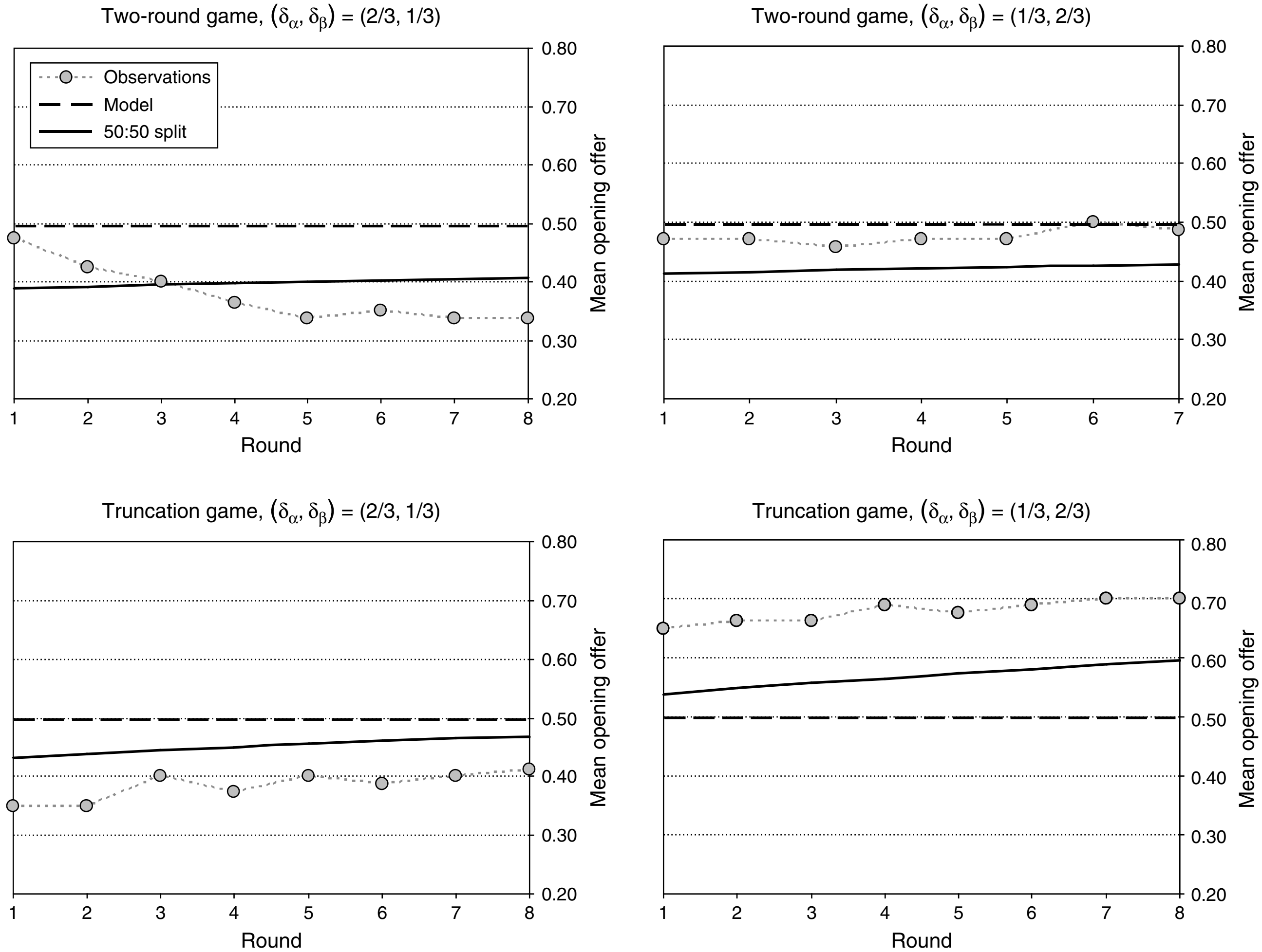
$$\text{Obs.} = \underset{(0.000)}{0.949}\ \text{Pred.} - \underset{(0.618)}{0.059}. \tag{10}$$

The intercept term is not significantly different from 0, the coefficient for PREDICTED is not statistically different from 1, and the model that forces the slope to 1 cannot be rejected (F-test, $p = 0.679$). This good fit, achieved despite considerable variance in the data, is particularly notable because the model was calibrated on a one-round version of the bargaining game, where counteroffers are impossible.

5.5. Experience Effects

In three studies (Neelin et al. 1988, Ochs and Roth 1989, Bolton 1991) game players repeated experience with bargaining in the same role. As an illustration, Figure 8 shows experience trends in the Bolton (1991) observations. In the (2/3, 1/3) treatment of the truncation game, the observed mean opening offers deviate from the selfish equilibrium in the direction of the

Figure 8 Mean Opening Offers by Round in Two-Round (Top) and Truncation (Bottom) Versions of Sequential Bargaining, with Different Discount Factors and a Pie of $12



Source. Bolton (1991).

equal money division. This 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 graphs for the Neelin et al. (1988) and Ochs and Roth (1989) data are comparable (and available in the online appendix).

The rate of change due to experience sometimes differs between the model and the data (i.e., the rate forecast by the model tends to be slower). Nevertheless, the model tends to capture the direction of the effect. In Table 2, we classify the slope parameters obtained from linear regressions on the data and model into three categories: positive trend (mean opening offers tend to increase with experience), not significant (at $p < 0.05$), and negative trend. Treatment data disperse essentially uniformly across these categories. For 8 of 13 experiments (61.5%), the experience effect predictions and observations fall into identical categories, which is significantly better than the proportion of success available through random selection (proportion test, one-tailed $p = 0.035$).

Another issue is whether, with experience, first offers move toward the (nonquantal) ERC equilibrium. In a simple test, we check whether the mean squared error of the observation versus the prediction tends to be smaller in the final than in the first round of play. Across 13 treatments, the average error decreases (one-tailed $p = 0.027$), which indicates that, with experience, offer behavior moves toward the ERC equilibrium.

5.6. Empirical Regularities

In Table 3, we summarize the major empirical regularities reported in bargaining game literature. Most are well predicted by the ERC model (Roth et al. 1991 data included).

6. Robustness Checks

In this section, we first report the results of three alternate specifications of the preference function: the simple quantal response equilibrium (QRE) model with

Table 2 Slope Parameters Obtained from Linear Regressions on Data and Model

Experiment	Initials	Rounds	Trend Obs.	Trend Model
Neelin et al. (1988)	NSS4	5	n.s.	+
Ochs and Roth (1989)	OR1	2	−	−
	OR2	2	n.s.	n.s.
	OR3	2	+	+
	OR4	2	−	+
	OR5	3	−	−
	OR6	3	−	−
	OR7	3	n.s.	n.s.
	OR8	3	+	−
Bolton (1991)	B1	2	−	+
	B2	2	n.s.	+
	B3	Trunc.	+	+
	B4	Trunc.	+	+
Convergent			8 of 13	0.615
Divergent			3 of 13	0.231

Notes. About 62% of slope predictions converge with the observations. See Table 1 for treatment label interpretations.

Table 3 Empirical Regularities in Bargaining Game Data

Empirical regularities	Data	Model	Accuracy (%)
Opening offers tend to be higher than the subgame perfect equilibrium	15 of 20	15 of 20	90
Opening offers tend to be less than half of the pie	19 of 20	19 of 20	100
Only exception to above: truncation game (0.33, 0.67)	✓	✓	
Rejection rates increase with discount factors (Güth and Tietz 1988, Ochs and Roth 1989)	9 of 11	11 of 11	82
A substantial proportion of rejected first-period offers are followed by disadvantageous counteroffers	17 of 19	19 of 19	89
Experience trends observed in mean opening offers	9 of 13	11 of 13	62

Notes. Most regularities are well predicted by the model. Accuracy is the proportion of model predictions that concur with observations (presence or absence of regularities).

no fairness considerations; the Fehr–Schmidt variant; and the ERC model, in which the social reference point for fairness is freely estimated. We also deal with another issue, namely, how dependent our estimates are on the particular data we use to fit the model. We fit the model in two alternative ways, using the two- and three-round data from one of the multiple-round studies. Finally, we extend our out-of-sample estimates to an experiment involving a three-person ultimatum game that the ERC model successfully explains in an ordinal sense (Bolton and Ockenfels 1998). This provides a strong test of stability with regard to the model's reference point; it also permits a comparison to a data set in which the proposer is far less generous than in any of the games examined thus far.

6.1. Alternative Fairness Specifications

In this section, we test simple QRE and Fehr–Schmidt specifications—and report both in-sample fit and out-of-sample predictions. To provide fair comparisons, we keep the learning components of the model unchanged and reestimate the parameters τ'_α, τ'_β, and τ_1 for each variant.

6.1.1. Simple QRE Model. The simple QRE model differs from our ERC model in that it assumes no fairness considerations; hence, the b parameter in the original ERC model is set to 0. This model assumes that bargainers are driven solely by pecuniary motives, though they make choices that are probabilistic in nature. Thus the model provides a test of the extent to which fairness considerations really are critical for explaining bargaining behavior. We fit this simpler model with the Roth et al. (1991) data and obtain the following parameter estimates (standard deviations in parentheses):

$$\begin{array}{cccc} b = \text{n/a} & \tau'_\alpha = 0.540 & \tau'_\beta = 0.389 & \tau_1 = 0.028 \\ (\text{n/a}) & (0.018) & (0.016) & (0.006) \end{array}.$$

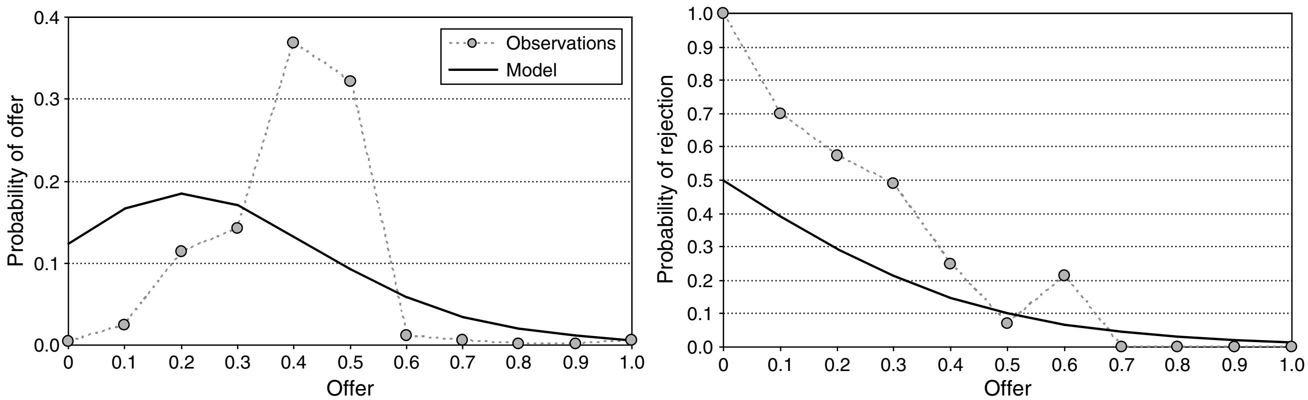
All parameters have the expected sign and are significant at $p < 0.01$. The model predicts that the average offer will be 29.5% of the pie (cf. 40.1% in the data), with an average rejection rate of 25.3% (cf. 26.4%). These summary statistics suggest that the QRE model provides a reasonable fit, but removing fairness considerations actually greatly deteriorates the fit, as shown in Figure 9.

If the proposer offers nothing to the responder ($\sigma = 0$), the QRE model predicts that the latter will be indifferent between accepting and rejecting the offer because both imply no payoff; the data clearly suggest that responders consider such offers profoundly unfair and reject them with near certainty. In the absence of fairness considerations, this pattern cannot be captured by the QRE model. Furthermore, the log-likelihood of the QRE model is $-3{,}311$ (standard deviation 22.9, bootstrap estimation), whereas its ERC counterpart is $-2{,}694$ (44.8). Thus, we can easily reject the former statistically. (Also see Yi 2005, who comes to similar conclusions from a different approach.)

6.1.2. Fehr–Schmidt Model. The Fehr–Schmidt model (Fehr and Schmidt 1999) is one of the important alternatives to ERC (a second, Charness and Rabin 2002, is omitted because it is not a good candidate to fit these data[7]). The Fehr–Schmidt model

[7] Charness and Rabin's model (2002, pp. 818–820) posits that Pareto-damaging games, such as ultimatum bargaining, involve a form of reciprocity called "concern withdrawal," such that players "withdraw their willingness to sacrifice to allocate the fair share toward somebody who himself is unwilling to sacrifice for the sake of fairness." The model they present captures this kind of

Figure 9 Probability That the Proposer Makes an Offer of σ_i (Left) and Probability That the Responder Rejects Such an Offer (Right), Observations vs. QRE Model



Source. Roth et al. (1991).

employs a more egalitarian measure of fairness in the preference function than does ERC. According to this specification, in a game with three or more players, two settlements that provide the same absolute payoff for one player might not lead to the same utility, depending on the fairness with which the rest of the pie gets split among the remaining players. In a two-player game, however, the utility function simplifies to a linear variant of the ERC model (with the assumption that $\alpha > \beta > 0$):

$$U(\sigma) = \begin{cases} c\left(\sigma - \alpha\left(\frac{1}{2} - \sigma\right)\right) & \text{if } \sigma < 1/2, \\ c\left(\sigma - \beta\left(\sigma - \frac{1}{2}\right)\right) & \text{if } \sigma > 1/2, \end{cases} \tag{11}$$

where c is the size of the pie, σ is the share of the pie offered, and α and β are negative and positive reciprocity parameters, respectively. We fit the Fehr–Schmidt model to the Roth et al. (1991) data and obtain the following results (standard deviation in parentheses):

$$\begin{array}{ccccc} \alpha = 1.053 & \beta = 0.003 & \tau'_\alpha = 0.877 & \tau'_\beta = 0.346 & \tau_1 = 0.040 \\ (0.076) & (0.007) & (0.108) & (0.025) & (0.013) \end{array}.$$

The $\beta > 0$ assumption of the model is violated, and the Fehr–Schmidt model reduces to a linear version of the original ERC, negative reciprocity specification. This result is consistent with our finding that positive reciprocity in the ERC model is mostly irrelevant. We cannot rule out positive reciprocity per se, but the ultimatum game to which we fit the model does not provide data that can quantify it.

The overall fit of the model is excellent (see Figure 10). The log-likelihood is $-2{,}586$ (45.8), compared with $-2{,}694$ (44.8) for the nonlinear ERC specification, which is a significant improvement (t-test $= 9.08$, $p < 0.01$). The correlations between observations and predictions are high, with $R_\alpha = 0.974$ and $R_\beta = 0.974$.

The positive reciprocity assumption of the Fehr–Schmidt model ($\beta > 0$) is violated by our estimations, and, because all games in our sample are two-person games, the only difference that remains with the ERC model is that the Fehr–Schmidt model specifies negative reciprocity as a linear function. As Fehr and Schmidt (1999, p. 823) argue, such linearity in inequality aversion "may not be fully realistic," but it is sufficient to provide a good fit in the one-round game. However, as Fehr and Schmidt (1999, p. 823) also report, "some observations on [other] experiments suggest that there are a nonnegligible fraction of people who exhibit nonlinear inequality aversion." For this reason, we might suspect that the nonlinear specification should provide better out-of-sample fit for more complex, multiple-round games.

In Table 4, we report the log-likelihood of our out-of-sample estimations for all multiple-round data sets, as derived from the ERC model (nonlinear inequality aversion), the Fehr–Schmidt model (linear inequality aversion), and the simple QRE model (no inequality aversion). The ERC model provides the best out-of-sample forecasts for 11 of the 18 games tested; the Fehr–Schmidt model forecasts best 4 of the 18 games and ties with ERC for 3 more. As expected, the QRE model fares the worst by far.

6.1.3. Alternative Fairness Reference Point. In the two-person game, the ERC model assumes a reference point of half of the pie; receiving less than that is deemed unfair and associated with a negative utility

reciprocity only "crudely," as they acknowledge (Charness and Rabin 2002, p. 825). Although they construct a more detailed version of the model in an appendix, they conclude that "It is too restrictive to be directly applied to experimental evidence…" (Charness and Rabin 2002, p. 857). The elaboration necessary to fit this model to the data is beyond the scope of this paper.

Table 4 Maximum Likelihood for Out-of-Sample Forecasts with the (a) ERC Model (Nonlinear Inequality Aversion), (b) Fehr–Schmidt Model (Linear Inequality Aversion), and (c) Simple QRE Model (No Inequality Aversion); All Calibrated with the Roth et al. (1991) Ultimatum Game Data Set

Experiment	Initials	Rounds	ERC	Fehr-Schmidt	QRE
Roth et al. (1991)	RPOZ	1	−2,694.0	−2,586.3	−3,311.6
Binmore et al. (1985)	BSS	2	Raw data not available		
Güth and Tietz (1988)	GT1	2	−207.0	−233.3	**−122.9**
	GT2	2	**−183.9**	**−186.0**	−209.3
Neelin et al. (1988)	NSS1	2	**−128.3**	−130.7	−135.2
	NSS2	3	−108.7	**−102.2**	−120.7
	NSS3	5	**−116.5**	**−118.1**	−126.1
	NSS4	5	**−170.7**	−191.2	−178.9
Ochs and Roth (1989)	OR1	2	−172.8	−239.4	**−156.0**
	OR2	2	−216.5	**−153.0**	−267.7
	OR3	2	**−137.4**	−162.1	−160.0
	OR4	2	−214.1	−222.4	**−198.7**
	OR5	3	−139.4	**−127.3**	−238.7
	OR6	3	**−211.4**	−247.6	−232.3
	OR7	3	**−203.4**	−241.7	−211.5
	OR8	3	**−247.3**	**−247.8**	−266.0
Bolton (1991)	B1	2	**−145.2**	−150.6	−182.3
	B2	2	−114.4	**−96.6**	−139.2
	B3	Trunc.	**−156.6**	−173.6	−189.5
	B4	Trunc.	**−144.9**	−157.7	−183.4
Best fit			11 of 18	4 of 18	3 of 18
n.s. ≠			0 of 7	3 of 12	0 of 15
Total			11 of 18	7 of 18	3 of 18

Notes. Bold, underlined values indicate the best fit. Bold, nonunderlined values indicate fits not statistically different from the best fit at $p < 0.05$. Shaded cells represent in-sample fit and are excluded from the comparisons.

component. This 50–50 reference point makes sense from a purely monetary perspective because beyond this point the share of the other player becomes more desirable than one's own.

To explore whether this assumption is justified, we modify the utility function in Equation (2) by replacing the constant 1/2 in the term $(\sigma - 1/2)$ with a parameter ω. When we freely estimate this additional parameter with the Roth et al. (1991) data, we determine $\omega = 0.521$ (standard deviation $= 0.087$), which is not significantly different from 1/2. Neither the overall fit of the model nor its out-of-sample predictive accuracy improves with any significance.

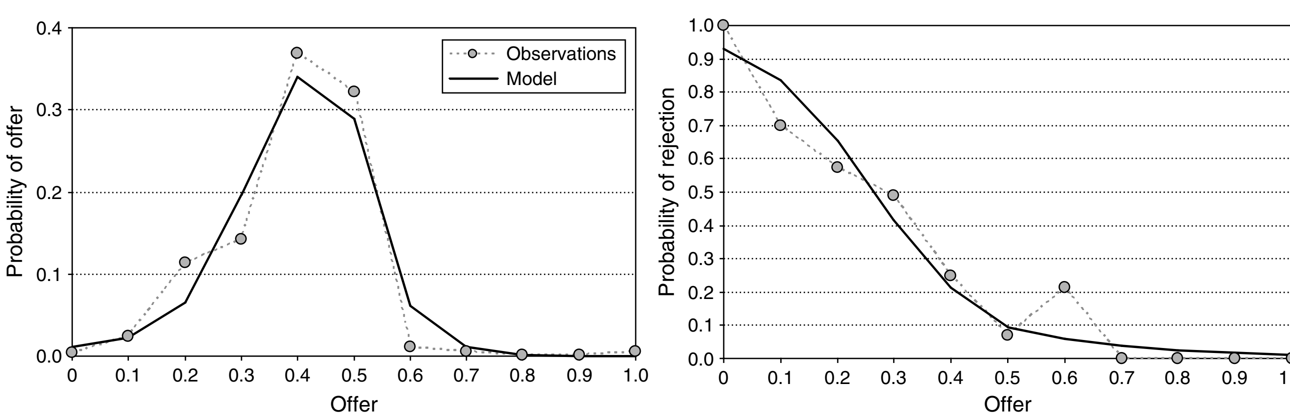
However, though this 50–50 social reference point works best in stylized games, it might not hold in real-life bargaining situations, in which bargaining powers vary and positional advantages exist. In these cases, the reference point should *not* be assumed but rather freely estimated.

6.2. Alternative Model Fits

A suitable data set for estimating the model needs two critical characteristics: First, to estimate experience effects, the data need to include observations of multiple plays. Second, because the desirable properties of maximum likelihood estimates can be achieved only asymptotically, the data set should be large. Of the data sets we examine, Ochs and Roth's (1989) meets both criteria best, with enough data to fit both two-round and three-round games separately—an interesting exercise considering the issue of backward induction's ability to approximate behavior. We reestimate the model parameters using the two- and three-round game data separately and estimate one set of parameters to fit all four treatments of each game (i.e., four combinations of discount factors) simultaneously, which creates two data sets of 380 observations each.

As we report in Table 5, the results of the parameter estimates suggest that most differences are statistically significant, though the b estimate does not differ statistically between the second and third data sets, nor does τ_1 (experience) between the first and

Figure 10 Probability That the Proposer Makes an Offer of σ_i (Left) and Probability That Such an Offer Is Rejected (Right), Observations vs. Fehr–Schmidt Model



Source. Roth et al. (1991).

Table 5 Parameter Estimates of the Model, Using Maximum Likelihood Estimation and Three Different Data Sets

	b	τ_α	τ_β	τ_1
Ultimatum game	6.692	0.690	0.280	0.065
(Roth et al. 1991)	(0.740)	(0.073)	(0.027)	(0.016)
Two-round bargaining game	10.066	0.586	0.222	0.007
(Ochs and Roth 1989)	(2.875)	(0.054)	(0.033)	(0.017)
Three-round bargaining game	12.081	0.502	0.191	0.068
(Ochs and Roth 1989)	(1.991)	(0.071)	(0.017)	(0.028)

third (i.e., learning does not accelerate as the number of rounds increases). In terms of behavioral tendencies, parameter estimates appear reasonably similar across games. Perhaps the biggest difference occurs with respect to the estimate of b, which increases with the number of rounds; the two-round and, to a greater extent, the three-round estimates imply that relative payoff takes on a somewhat heavier weight. The heavier weight might result from the greater difficulties associated with backward induction, resulting in decisions that add somewhat more weight on what is fair and less on strategic calculation.

We report the out-of-sample predictive accuracy of the ERC model in Table 6, based on the data sets on which we calibrate the model. Specifically, for the ERC model calibration, we use the simplest version of the ultimatum game, which best predicts the Neelin et al. (1988) and Bolton (1991) data sets. In contrast, the model calibrated on the Ochs and Roth (1989) data predicts other games best. Although the one-round game seems to provide the best basis for calibration, differences in log-likelihoods are small; therefore, the specific data set on which a model is calibrated has only a marginal impact on out-of-sample predictive accuracy.

6.2.1. Alternative Model Fits with the Fehr–Schmidt Model. The Fehr–Schmidt model provides an excellent fit to the one-round ultimatum data set but is outperformed by the ERC model when it comes to out-of-sample forecasts. We also test whether its poor out-of-sample performance may be due to the data set on which it was fit by estimating it on Ochs and Roth's (1989) two- and three-round data sets and predicting the other games out-of-sample (as for the ERC model in the previous section; see Table 7 for details).

When we fit both the ERC and the Fehr–Schmidt models on Ochs and Roth's (1989) two-round data sets, the latter offer better in-sample fit but poorer out-of-sample performance, with particularly bad performance in terms of predicting the one-round ultimatum game. If both models use the three-round bargaining data sets, their in-sample and out-of-sample results are comparable, except that the Fehr–Schmidt model again forecasts the one-round ultimatum game

Table 6 Maximum Likelihood of Out-of-Sample Forecasts of the ERC Model, Calibrated on the (a) Roth et al. (1991) Ultimatum Game Data Set, (b) Ochs and Roth's (1989) Two-Round Game Data Sets, and (c) Ochs and Roth's (1989) Three-Round Game Data Sets

			ERC model fit on...		
Experiment	Initials	Rounds	RPOZ	OR1–OR4	OR5–OR8
Roth et al. (1991)	RPOZ	1	−2,693.96	2,833.3	**−2,777.9**
Binmore et al. (1985)	BSS	2	Raw data not available		
Güth and Tietz (1988)	GT1	2	−207.0	**−195.7**	**−199.6**
	GT2	2	−183.9	**−175.2**	**−174.1**
Neelin et al. (1988)	NSS1	2	**−128.3**	−131.3	−132.0
	NSS2	3	**−108.7**	**−110.9**	**−110.1**
	NSS3	5	**−116.5**	−119.8	−120.1
	NSS4	5	**−170.7**	−178.0	−180.4
Ochs and Roth (1989)	OR1	2	**−172.8**	−159.6	**−174.4**
	OR2	2	−216.5	−177.0	**−170.1**
	OR3	2	**−137.4**	−141.3	**−136.5**
	OR4	2	−214.1	−192.7	**−200.8**
	OR5	3	**−139.4**	**−140.4**	−134.2
	OR6	3	−211.4	**−206.1**	−204.9
	OR7	3	**−203.4**	**−201.4**	−199.5
	OR8	3	−247.3	**−220.6**	−221.6
Bolton (1991)	B1	2	**−145.2**	−158.0	−156.4
	B2	2	**−114.4**	−123.9	−120.3
	B3	Trunc.	**−156.6**	−167.6	−164.9
	B4	Trunc.	**−144.9**	−178.8	−171.9
Best fit			10 of 18	4 of 13	5 of 13
n.s. ≠			2 of 8	3 of 9	3 of 8
Total			12 of 18	7 of 15	8 of 15

Notes. Bold, underlined values indicate the best out-of-sample fit. Bold, nonunderlined values indicate fit not statistically different from the best fit at $p < 0.05$. Shaded cells represent in-sample fit and are excluded from comparisons.

poorly out of sample. To provide a holistic view, we sum the log-likelihoods for all out-of-sample forecasts; regardless of the data set used, the ERC model provides a better out-of-sample fit at $p < 0.05$.

The linear specification of the Fehr–Schmidt models provides an excellent fit (sometimes better than the ERC model). However, when the strategic parameters of the game push the negotiation solutions outside of the boundaries of the data set used to fit the model, the linear specification cannot anticipate the wide variations observed in many games. The ERC specification therefore offers better out-of-sample predictive power, regardless of the data set used.

6.3. 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 a third player ("dummy") also

Table 7 **Maximum Likelihood of Out-of-Sample Forecasts of the ERC and Fehr–Schmidt Models, Calibrated on the (a) Roth et al. (1991) Ultimatum Game Data Set, (b) Ochs and Roth's (1989) Two-Round Game Data Sets, and (c) Ochs and Roth's (1989) Three-Round Game Data Sets**

			RPOZ data		OR1–OR4 data		OR5–OR8 data	
Experiment	Initials	Rounds	ERC	Fehr–Schmidt	ERC	Fehr–Schmidt	ERC	Fehr–Schmidt
Roth et al. (1991)	RPOZ	1	−2,694.0	**−2,586.3**	**−2,833.3**	−3,087.8	**−2,777.9**	−3,025.4
Binmore et al. (1985)	BSS	2			Raw data not available			
Güth and Tietz (1988)	GT1	2	**−207.0**	−233.3	−195.7	**−185.9**	−199.6	**−176.4**
	GT2	2	**−183.9**	−186.0	**−175.2**	−179.7	**−174.1**	−176.8
Neelin et al. (1988)	NSS1	2	**−128.3**	−130.7	**−131.3**	−131.7	**−132.0**	−133.7
	NSS2	3	−108.7	**−102.2**	**−110.9**	−115.8	**−110.1**	−113.9
	NSS3	5	**−116.5**	−118.1	**−119.8**	−122.9	**−120.1**	−123.9
	NSS4	5	**−170.7**	−191.2	**−178.0**	−178.3	−180.4	**−178.2**
Ochs and Roth (1989)	OR1	2	**−172.8**	−239.4	**−159.6**	−165.9	−174.4	**−161.7**
	OR2	2	−216.5	**−153.0**	−177.0	**−158.4**	−170.1	**−162.3**
	OR3	2	**−137.4**	−162.1	−141.3	**−124.2**	−136.5	**−125.7**
	OR4	2	**−214.1**	−222.4	−192.7	**−174.3**	−200.8	**−178.3**
	OR5	3	−139.4	**−127.3**	**−140.4**	−141.9	**−134.2**	−135.5
	OR6	3	**−211.4**	−247.6	**−206.1**	−210.6	**−204.9**	−208.5
	OR7	3	**−203.4**	−241.7	**−201.4**	−205.1	**−199.5**	−200.9
	OR8	3	**−247.3**	−247.8	−220.6	**−214.3**	−221.6	**−219.2**
Bolton (1991)	B1	2	**−145.2**	−150.6	**−158.0**	−162.0	**−156.4**	−160.1
	B2	2	−114.4	**−96.6**	−123.9	**−119.8**	−120.3	**−117.6**
	B3	Trunc.	**−156.6**	−173.6	−167.6	**−167.0**	**−164.9**	−169.4
	B4	Trunc.	**−144.9**	−157.7	**−178.8**	−195.2	**−171.9**	−186.2
Best fit			10 of 18	4 of 18	6 of 15	3 of 15	6 of 18	5 of 18
n.s. ≠			4 of 8	0 of 14	5 of 9	1 of 12	2 of 12	2 of 13
Total			14 of 18	4 of 18	11 of 15	4 of 15	8 of 18	7 of 18
Global log-likelihood (out of sample)			**−3,018.5**	−3,181.2	**−5,140.9**	−5,418.0	**−4,989.6**	−5,189.6

Notes. Bold, underlined values indicate the best out-of-sample fit. Bold, nonunderlined values indicate fit not statistically different from the best fit at $p < 0.05$. Shaded cells represent in-sample fit and are excluded from comparisons.

takes a share of the pie, though this player makes no decisions. The first player proposes a division of the pie among all three players, and the second player either accepts or rejects that offer. If the proposition is rejected, all players receive no monetary payoff. We compare our predictions with the simplest version (i.e., essential information treatment, constant mode) of the original experiment conducted by Güth and van Damme (1998), in which players had to share a pie of 24 Dutch guilders (divided into 120 tokens), which represented approximately \$13.6 at that time ($c = 13.6$).[8]

ERC stipulates a modification in the utility function to fit a three-person game: Because three 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 (2), the term $(\sigma - 1/2)$ gets replaced by $(\sigma - 1/3)$ (Bolton and Ockenfels 1998 study this game using the same modification). This is the only change: 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. (1991) ultimatum game with a social reference share of one-half.

One of Güth and van Dammes's (1998) critical findings was that the dummy player receives little more than the minimum, five tokens, that the experimenters required. On average, during the first six games, the dummy's share was 7.8 of 120 tokens in the observations (6.5%). In contrast, our model predicts 17.2 (14.3%), as shown in Table 8.

A second finding was that the rejection rates in this game were smaller than they tend to be in regular ultimatum games. The average rejection rate in the Roth et al. (1991) two-person ultimatum experiment is 0.264 (the typical 15%–20% rejection rate observed in two-person ultimatum games (Roth 1995)). However, the rate is 0.097 in Güth and van Damme's (1998) data

[8] Güth and van Damme (1998) also report data pertaining to when the responder knows the entire proposed allocation. These data are very similar to the data we report.

Table 8 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 Güth and van Damme's (1998) Game, Observations vs. Model

	Observations, Games 1–6	Model		
		Games 1–6	15th game	30th game
Proposer (x)	79.1	65.7	72.1	78.0
Responder (y)	33.1	37.1	36.7	33.8
Dummy (z)	7.8	17.2	11.1	8.2
Rejection rate	0.097	0.256	0.152	0.088

Source. Güth and van Damme (1998).
Note. The model replicates observations when players gain experience: The dummy's payoff decreases and the proposer's payoff increases with experience.

set (i.e., essential information condition). Our model does not capture this finding, predicting a rejection rate of 0.256. The problem appears to be that real players learn more quickly than our simulated players; the predicted rejection rate after 30 games is 0.088.

There was also an experience effect in the data, and our model does capture some of its characteristics. Specifically, we apply the model by predicting the game's outcomes for the sixth, 15th, and 30th games. As shown in Table 8, this does not affect the responder's payoff much, but it 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. Güth and van Damme (1998) identify the same pattern in their experiment.

Thus, the model correctly estimates both the nature and the direction of change due to player experience, but it underestimates the pace at which that change occurs. The model's predictions for the 30th game come strikingly close to the observations made during the sixth game in Güth and van Damme's (1998) experiment, namely, $x = 80.8$, $y = 33.3$, and $z = 5.8$, with a similar rejection rate. Thus, the large overestimation of the rejection rate mainly results from the model's inability to predict the pace of learning rather than any failures in predicting the direction, nature, or effects of such learning.

7. Summary

We estimate fairness preferences from the simplest, one-round version of sequential bargaining games and use these estimates to forecast, out of sample, multiple-round games with various lengths, discount factors, pie sizes, and levels of bargainer experience. The out-of-sample forecasts capture many of the reported empirical regularities, as well as a substantial amount of the variability in first offers, rejections, and counteroffers (even though there are no counteroffers in the one-round version used to fit the model). Overall, they offer better predictions than traditional preference models that ignore fairness considerations. In statistical tests in which we compare the forecasts to actual data, we cannot reject the model, with the exception of the rejection rates for the smallest pie size games. In this sense, willingness to pay for fair treatment is robust across the seven studies we examine, and much of the variance in observed behavior can be accounted for by changes in the potential trade-offs, which themselves result from changes in the strategic parameters across games. Also, experience effects tend to push offers in the direction of the stationary ERC equilibrium—a point potentially important for future theory pertaining to learning and bargaining. Finally, our model compares well with alternative preference specifications, and the estimates are reasonably robust to alternative data fittings. These findings lay the groundwork for estimating the influence of fairness on field negotiations and across different subpopulations. However, extending our work to these tasks entails certain challenges:

For example, obtaining estimates for field negotiations from the present model would require priming the model with the fairness standard that exists in the field. The key fairness standard among the bargainers in these studies was how their share compares with that of the other bargainer(s). But the comparison also might be calibrated to those aspects relevant in other negotiations, such as the outside (pattern) agreements that often influence collective bargaining settlements. An existing body of work identifies some fairness criteria; for example, Young (1991) explores the issues involved in choosing between competing fairness criteria.

It also would be interesting to investigate whether different social strata exhibit similar or different attitudes toward fairness. The subjects in these studies are all university students, so the method of estimation we present should be extended to different groups of respondents. Although we find little evidence to suggest any significant cultural effects, an investigation focused specifically on this issue might identify effects not evident here. Moreover, some social status or life cycle effects seem likely, which could have great importance for negotiation practitioners.

Heterogeneity also could mark attitudes among people in the same social strata. We are limited to estimating average fairness preferences, whereas addressing individual differences would require refining the way in which the model accounts for individual variability. Players' heterogeneity and individuals' choice randomness (two main explanations of why games outcomes are probabilistic) become somewhat confounded in our formulation, particularly in the parameter of the decision rules, and cannot be analyzed or estimated separately. The major obstacle to overcoming this difficulty is getting data sets that are large enough, in that they provide a sufficient number of choices per individual. Alternatively, Bayesian statistical methods might be brought into play.

Also, the experience component of the ERC utility decision framework is independent of bargainers' past actions, which seems rather unrealistic. The experience trend embedded in the model supposedly arises from a decrease in bargainers' heterogeneity and choice randomness, but the way these phenomena relate to experience and trial and error requires additional investigation.

Finally, it would be useful to extend our analysis to bargaining protocols beyond the offer—counteroffer format. In this regard, the ERC model implies some interesting directions. For example, introducing competition into the negotiation may lead to competitive outcomes, even if they are highly inequitable, despite any preferences for fairness. Failure to compete leaves one player with nothing and others with much—a bad outcome in both absolute and relative terms. In contrast, competing at least fulfills a player's absolute objectives. Roth et al. (1991) include a market game modification of the ultimatum game, in which multiple proposers make offers to a single responder, who could choose to accept just one proposal. Whether players deem the procedures followed in the bargaining protocol fair also might have an impact. Pratt and Zeckhauser (1990) describe the use of a fair procedure for allocating nonmonetary items; Brams and Taylor (1993) describe a generalized version of the "one divides, the other chooses" protocol for fair division; and Bolton et al. (2005) sketch how the ERC model might be extended to account for fair procedures. We hope to have something to say about these kinds of protocols in the near future.

8. Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at http://mansci.journal.informs.org/.

Acknowledgments

This manuscript was part of the first author's doctoral thesis at Pennsylvania State University. The authors thank Gary Lilien and Axel Ockenfels for careful reviews of previous versions of the manuscript. The second author gratefully acknowledges the financial support of the National Science Foundation.

Appendix. Opening Offer and Rejection Behavior by Treatment: Observations and Model Predictions (ERC, Fehr–Schmidt, and QRE Models)

		Opening offers				Rejection rates				Disadvantageous counteroffers			
Experiment	Initials	Obs.	ERC	Fehr–Schmidt	QRE	Obs.	ERC	Fehr–Schmidt	QRE	Obs.	ERC	Fehr–Schmidt	QRE
Roth et al. (1991)	RPOZ	0.395	0.393	0.401	0.289	0.272	0.265	0.258	0.259			n/a	
Binmore et al. (1985)	BSS	0.416	0.441	0.441	0.441	0.148	0.488	0.488	0.488	0.750	0.788	0.788	0.788
Güth and Tietz (1988)	GT1	0.276	0.373	0.391	0.292	0.190	0.288	0.268	0.246	0.750	1.000	1.000	0.950
	GT2	0.440	0.409	0.398	0.446	0.619	0.572	0.588	0.565	0.000	0.108	0.089	0.058
Neelin et al. (1988)	NSS1	0.265	0.355	0.350	0.377	0.225	0.463	0.460	0.370	0.556	0.663	0.645	0.634
	NSS2	0.472	0.354	0.349	0.380	0.050	0.499	0.505	0.416	0.500	0.395	0.367	0.434
	NSS3	0.320	0.355	0.349	0.378	0.125	0.476	0.476	0.386	0.400	0.566	0.546	0.561
	NSS4	0.348	0.436	0.450	0.365	0.167	0.326	0.308	0.329	0.857	0.849	0.869	0.448
Ochs and Roth (1989)	OR1	0.399	0.418	0.442	0.372	0.100	0.103	0.115	0.153	0.600	0.999	0.998	0.579
	OR2	0.482	0.425	0.464	0.382	0.150	0.139	0.120	0.175	1.000	0.999	0.998	0.620
	OR3	0.471	0.466	0.495	0.487	0.187	0.207	0.094	0.223	0.733	0.773	0.969	0.450
	OR4	0.458	0.441	0.482	0.469	0.200	0.159	0.090	0.186	0.550	0.977	0.988	0.355
	OR5	0.429	0.414	0.429	0.338	0.120	0.096	0.108	0.153	1.000	0.956	0.940	0.546
	OR6	0.443	0.433	0.462	0.389	0.140	0.303	0.237	0.307	0.857	0.764	0.691	0.559
	OR7	0.449	0.434	0.462	0.395	0.144	0.314	0.242	0.321	0.462	0.341	0.348	0.225
	OR8	0.453	0.425	0.450	0.401	0.289	0.135	0.129	0.176	0.885	0.928	0.888	0.296
Bolton (1991)	B1	0.378	0.398	0.413	0.290	0.188	0.320	0.302	0.352	0.833	0.848	0.817	0.462
	B2	0.476	0.421	0.435	0.391	0.204	0.322	0.291	0.386	0.200	0.471	0.474	0.199
	B3	0.384	0.452	0.464	0.340	0.391	0.365	0.341	0.403	0.960	0.885	0.871	0.493
	B4	0.678	0.568	0.547	0.491	0.266	0.539	0.569	0.570	0.000	0.344	0.259	0.230
Güth and van Damme (1998)[a]	GvD	0.276	0.309	0.319	0.284	0.097	0.256	0.240	0.259			n/a	
		0.065	0.143	0.148	0.131			n/a					
Pearson's R^2 (All observations)			0.789	0.739	0.654		0.118	0.120	0.198		0.694	0.676	0.347
(Excluding RPOZ and GvD)			0.596	0.495	0.416		0.120	0.122	0.207		0.694	0.676	0.347
(Excluding RPOZ, GvD, and small pie games)			0.707	0.612	0.562		0.447	0.468	0.444		0.705	0.686	0.454

[a]The top number refers to mean offer to the responder, and the bottom number refers to mean offer to the dummy.

References

Binmore, K., A. Shaked, J. Sutton. 1985. Testing noncooperative bargaining theory: A preliminary study. *Amer. Econom. Rev.* **75**(December), 1178–1180.

Boiney, L. G. 1995. When efficient is insufficient: Fairness in decisions affecting a group. *Management Sci.* **41**(9) 1523–1537.

Bolton, G. E. 1991. A comparative model of bargaining: Theory and evidence. *Amer. Econom. Rev.* **81**(5) 1096–1136.

Bolton, G. E., A. Ockenfels. 1998. Strategy and equity: An ERC-analysis of the Güth-van Damme game. *J. Math. Psych.* **42**(2/3) 215–226.

Bolton, G. E., A. Ockenfels. 2000. ERC: A theory of equity, reciprocity, and competition. *Amer. Econom. Rev.* **90**(1) 166–193.

Bolton, G. E., J. Brandts, A. Ockenfels. 2005. Fair procedures: Evidence from games involving lotteries. *Econom. J.* **115** 1054–1076.

Brams, S. J., A. D. Taylor. 1993. An envy-free cake division protocol. *Amer. Math. Monthly* **102** 9–18.

Charness, G., M. Rabin. 2002. Understanding social preferences with simple tests. *Quart. J. Econom.* **117**(3) 817–869.

Costa-Gomes, M., K. G. Zauner. 2001. Ultimatum bargaining behavior in Israel, Japan, Slovenia, and the United States: A social utility analysis. *Games Econom. Behav.* **34** 238–270.

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

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

Fehr, E., K. Schmidt. 1999. A theory of fairness, competition, and cooperation. *Quart. J. Econom.* **114** 817–868.

Forsythe, R., J. Horowitz, N. E. Savin, M. Sefton. 1994. Fairness in simple bargaining experiments. *Games Econom. Behav.* **6**(3) 347–369.

Gächter, S., A. Riedl. 2005. Moral property rights in bargaining with infeasible claims. *Management Sci.* **51**(2) 249–263.

Goeree, J. K., C. A. Holt. 2000. Asymmetric inequality aversion and noisy behavior in alternating-offer bargaining games. *Eur. Econom. Rev.* **44** 1079–1089.

Güth, W., R. Tietz. 1988. Ultimatum bargaining for a shrinking cake—An experimental analysis. R. Tietz, W. Albers, R. Selten, eds. *Bounded Rational Behavior in Experimental Games and Markets*. Springer, Berlin, 111–128.

Güth, W., E. van Damme. 1998. Information, strategic behavior and fairness in ultimatum bargaining, an experimental study. *J. Math. Psych.* **42**(2/3) 227–247.

Luce, D. R. 1959. *Individual Choice Behaviour*. Wesley, New York.

McKelvey, R. D., T. R. Palfrey. 1995. Quantal response equilibria for normal-form games. *Games Econom. Behav.* **7** 6–38.

Narasimhan, C., C. He, E. Anderson, L. Brenner, P. Desai, D. Kuksov, P. Messinger, S. Moorthy, J. Nunes, Y. Rottenstreich, R. Staelin, G. Wu, Z. J. Zhang. 2005. Incorporating behavioral anomalies in strategic models. *Marketing Lett.* **16** 361–373.

Neelin, J., H. Sonnenschein, M. Spiegel. 1988. A further test of noncooperative bargaining theory. *Amer. Econom. Rev.* **78**(September) 824–836.

Ochs, J., A. E. Roth. 1989. An experimental study of sequential bargaining. *Amer. Econom. Rev.* **79**(June) 355–384.

Pratt, J. W., R. J. Zeckhauser. 1990. The fair and efficient division of the Winsor family silver. *Management Sci.* **36**(11) 1293–1301.

Rapoport, A., I. Erev, R. Zwick. 1995. An experimental study of buyer-seller negotiation with one-sided incomplete information and time discounting. *Management Sci.* **41**(3) 377–394.

Roth, A. E. 1995. Bargaining experiments. J. Kagel, A. E. Roth, eds. *Handbook of Experimental Economics*. Princeton University Press, Princeton, NJ, 256–269.

Roth, A. E., I. Erev. 1995. Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games Econom. Behav.* **8**(Special Issue: Nobel Symposium) 164–212.

Roth, A. E., V. Prasnikar, M. Okuno-Fujiwara, S. Zamir. 1991. Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An experimental study. *Amer. Econom. Rev.* **81**(5) 1068–1095.

Samuelson, W. 1980. First-offer bargains. *Management Sci.* **26**(2) 155–164.

Yi, K. 2005. Quantal-response equilibrium models of the ultimatum bargaining game. *Games Econom. Behav.* **51**(2) 324–348.

Young, H. P. 1991. Fair division. H. P. Young, ed. *Negotiation Analysis*. University of Michigan Press, Ann Arbor, 25–46.

Zwick, R., X.-P. Chen. 1999. What price fairness? A bargaining study. *Management Sci.* **45**(6) 804–823.