View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Research Papers in Economics

Asymmetric Inequality Aversion and Noisy Behavior in Alternating-Offer Bargaining Games

Jacob K. Goeree and Charles A. Holt[*]

Department of Economics, Rouss Hall, University of Virginia, Charlottesville, VA 22903

In two-stage bargaining games with alternating offers, the amount of the pie that remains after a rejection is what the first player should offer to the second player, since the second player can capture this remainder in the final (ultimatum) stage. Fairness considerations will reduce the correlation between first-stage offers and the size of the remaining pie, but randomness in behavior will have the same "flattening" effect. This paper reports an experiment designed to separate these considerations, by introducing asymmetric fixed money payments to each player. These endowments do not affect the perfect positive correlation between initial Nash offers and the remaining pie, but are selected to induce a perfectly *negative* relationship between the remaining pie size and the first-stage offer that would equalize final earnings of the two players. This negative relationship is apparent in the data, which suggests the importance of fairness considerations. A theoretical model of asymmetric inequality aversion and stochastic choice is used to provide maximum likelihood estimates of utility and logit error parameters. The parameters representing "envy," "guilt," and logit errors are all significant, and the resulting model produces the observed negative relationship between initial offers and residual pie size.

1. Introduction

Alternating-offer bargaining games have interested theorists because powerful backward induction arguments can be used to select a unique outcome from the wide range of Nash equilibria. Experimentalists in economics and psychology have been curious about these sharp predictions that require considerable amounts of common knowledge, strategic rationality, and willingness to accept unequal divisions. This paper reports a laboratory experiment designed to highlight the conflict between fairness and strategic considerations.

In a standard alternating-offer bargaining game, one player proposes a division of an amount of money that the other player can either accept or reject. If the proposer's offer is rejected, then the amount to be divided is reduced to some level denoted by R, and the responder

[*] This research was funded in part by the National Science Foundation (SBR-9818683). We wish to thank Ellen Quarrels and John Turner for helpful suggestions.

makes a counter proposal of how to split this residual. Earnings for both players are zero if the counter proposal is rejected. In a subgame perfect Nash equilibrium, the amount R that would remain after a rejection is what the first player should offer to the second player, since this is what the second player can capture in the final (ultimatum) stage.[1] Therefore, the Nash offers will be perfectly correlated with R.

Previous experimental evidence has shown that the correlation is generally positive, but less than perfect: proposers offer substantial amounts even when R is close to zero, and they do not offer anything close to the full pie amount when R equals the original amount to be divided (Roth, 1995). One possible explanation is that participants are to some extent concerned with relative earnings; so proposers are averse to giving up large fractions of the pie, or are hesitant to demand a high fraction knowing that responders may reject unfavorable splits, even when such splits are consistent with equilibrium predictions. This perspective is supported by the observation of "disadvantageous counteroffers" reported by Ochs and Roth (1989), who observe that rejections are often followed by counter-proposals that demand even less than the offer just rejected. As Bolton (1991) notes, this seemingly irrational behavior could be explained if players are motivated both by their own earnings and by relative earnings. "Irrational" rejections can also be explained by reactions to offers that are seen as being motivated by unkind or unfriendly intentions (Rabin, 1993), or by concerns for equity and inequality aversion (Bolton and Ockenfels, 1999; Fehr and Schmidt, 1999). Our experiment will allow an independent evaluation of some of these theories, which are discussed in more detail in section 3 below.

An alternative explanation of the less than perfect correlation between initial offers and the residual R is that unobserved factors and "noise" may make it hard to predict the reaction of a player who is almost indifferent between accepting or rejecting a small offer. As shown below, noise in decision making reduces the correlation between the initial offer and the amount R remaining in the second stage, which provides another explanation of observed data patterns in two-stage bargaining games.

Theories of noisy behavior that rely only on the costs of "errors" will be unaffected by

[1] To see this, note that the proposer "should" accept a one penny counter offer in the second stage, since a rejected counter offer will result in a zero payoff. Therefore, the responder can earn R minus a penny by rejecting the initial offer, so an initial offer of R is the lowest offer that will be accepted.

additional fixed payments that are received independent of the bargaining process. Our approach is to introduce asymmetric fixed payments that accentuate earnings inequities arising in a subgame perfect Nash equilibrium. The resulting data provide a platform for the development and refinement of models of noisy behavior that is affected by equity considerations.

The procedures and treatments are described in section 2, and the data patterns are summarized in section 3, with reference to alternative theoretical explanations of bargaining behavior. The fourth section presents a formal model of inequality aversion and stochastic choice. This model incorporates the insights of Bolton and Ockenfels (1999) and Fehr and Schmidt (1999) into an a logit equilibrium analysis motivated by McKelvey and Palfrey (1995). The final section concludes.

2. Experimental Design and Procedures

The two-stage bargaining game involves an initial pie of size \$2.40 in all treatments, which are parameterized by the amount of the pie remaining, as shown in the top row of Table 1. In addition to earnings from this bargaining process, each player receives an additional fixed payment that is independent of the bargaining outcome, and hence has no effect on the subgame perfect Nash prediction. The fixed payments for the initial proposer and the responder depend on the treatment, as shown in second and third rows of Table 1. Notice that the payment to the responder in the bottom row is equal to the remainder, R, in the top row, plus a constant of twenty-five cents.

Recall that the proposer need only offer an amount R to the responder in equilibrium, so the responder earns a low fixed payment precisely in the treatments for which the responder has low earnings in the subgame perfect Nash equilibrium. Thus the fixed payments exaggerate the earnings inequality arising in a subgame perfect Nash equilibrium, which will create a conflict between strategic and fairness considerations. In fact, the fixed payments in Table 1 were chosen so that equal final earnings would require a perfect negative relationship between the remaining pie size and first-stage offer. For example, when $R = 0$, the fixed payments are \$2.65 for the proposer and \$0.25 for the responder, so the offer that equalizes earnings is \$2.40. Conversely, when R is large the responder's fixed payment is high and a low offer equalizes earnings. To summarize, the subgame perfect Nash prediction involves a perfect positive correlation between

Table 1. Payoff Structure

	parameters for the seven bargaining games						
remaining pie size ($R)	0.00	0.40	0.80	1.20	1.60	2.00	2.40
fixed payment to the initial proposer	2.65	2.25	1.85	1.45	1.05	0.65	0.25
fixed payment to the initial responder	0.25	0.65	1.05	1.45	1.85	2.25	2.65

initial offers and R, and the equalization of final earnings produces a perfect negative correlation. Even if these relationships were "flattened" by noisy behavior, the design makes it possible to detect fairness considerations that tend to produce a negative correlation.

The subjects in the experiment were recruited at the University of Virginia. Participants were paid $5.00 for showing up and were told that they would be paid all earnings in cash, privately, immediately after the session, and that there would be no group discussion of earnings. The experiment consisted of six one-hour sessions, each with a different group of eight subjects, except as noted below.

Subjects were seated in private booths upon arrival, and instructions (available from the authors on request) were read aloud. Half of the participants were "type A" players who would make initial proposals to divide the $2.40. The proposal sheet used to communicate offers and counter offers clearly specified the fixed payment for A, the fixed payment for B, and the residual R, so these were common knowledge. The type A participants had to make 7 proposals, corresponding to the seven columns of Table 1. There was a separate proposal sheet for each proposal, and the 7 sheets for each type A person were given to that person in a random order. The initial proposed division of the $2.40 for the first sheet had to be recorded and returned before the second sheet was provided, so the type A people did not know the menu of proposal situations in advance. The type A people made all of their decisions without receiving feedback from the type B people. After the 7 proposed divisions had been made, numbered ping pong balls were used to assign each proposal to a type B player.[2] The type B participants received the randomly ordered sheets one at a time and made their responses, either accepting the division

[2] There were only 3 proposers and 3 responders in one session, and in one session there were 4 proposers and 3 responders, with one responder making two responses in each round. The exact matchings are shown in the data appendix available from the authors on request.

or rejecting and returning a counter proposal to divide the residual. Any rejections were returned to the matched type A for that proposal sheet, who would accept or reject. Thus each person earned 7 division amounts and 7 fixed payments, and the total earnings were computed and presented to subjects after they had completed a second, unrelated experiment (a series of six-stage centipede games). Earnings ranged from about $20 to $40 for the two-hour session.

3. Data Patterns and Implications

The average first-period offers are shown in Figure 1, with the fraction of the pie remaining on the horizontal axis and the proposer's offer (as a fraction of $2.40) on the vertical axis.[3] One apparent feature of the data is the abundance of equal divisions when the fixed payments for the proposers and responders are the same and the fraction of the pie remaining is one-half. This is not surprising since the Nash equilibrium offer of $1.20 coincides with a "fair" outcome with equal final payoffs for proposers and responders.[4]

The most salient feature of Figure 1, however, is the reversal in the predicted equilibrium relationship between the size of the pie remaining and the offer made. Recall that in a Nash equilibrium the proposer's offer equals the pie remaining since this is what the responder can request in the second-stage ultimatum game. This perfect linear relationship is generally not observed in two-stage bargaining experiments without fixed payments, but offers are positively correlated with the size of the pie remaining (e.g. Davis and Holt, 1993, chapter 5). Figure 1 shows that with asymmetric fixed payments that exacerbate the proposer's (dis)advantageous position, initial offers *decline* with the size of the pie remaining. This fixed-payment effect indicates that (strategic) fairness considerations play an important role.

Given the prevalence of "fair" proposals that tend to equalize earnings, it is not surprising that a fairly high percentage (75%) of the initial proposals were accepted. Rejections result in a final-stage ultimatum game, where the type B players on average offered about 1/3 of the

[3] A complete data appendix is available from the authors on request.

[4] Close to 80 percent of the type A subjects proposed an equal division in this case, resulting in an average offer of $1.17 (with a standard deviation of $0.09). Interestingly, the responders reacted most strongly to asymmetric or "unfair" offers in this symmetric case: offers that differed more than ten cents from an equal division were all rejected.

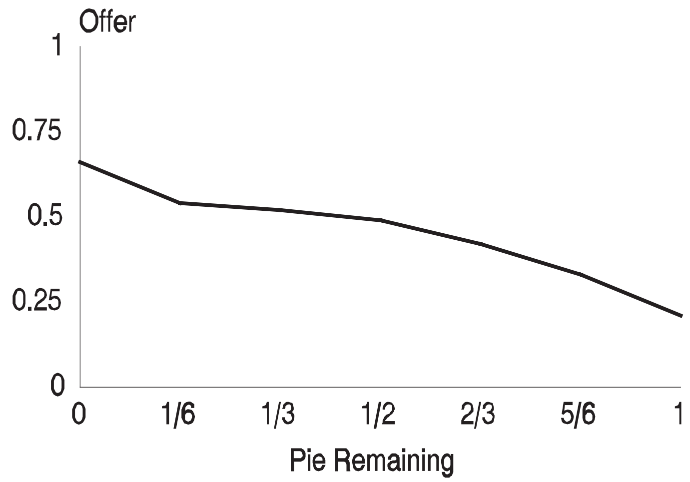

Figure 1. Average First-Stage Offers (dark line) and Standard Deviations (thin lines)

remaining pie to the type A players. Other aspects of the data will be discussed in the next section in the context of a formal model of asymmetric inequality aversion and noisy decision making. In order to motivate this model, we review some theories that have previously been successfully applied to explain data from bargaining and other experiments.

First, consider Gary Bolton's (1991) "comparative payoff" model in which a player's utility is increasing in both monetary and relative payoffs. The assumed relative payoff effect is asymmetric: when the ratio of a player's payoff to the other's payoff is less than one, the player's utility is increasing in this ratio, while it is independent of this ratio otherwise. This model provides a theoretical explanation for so called "disadvantageous counteroffers," i.e. counter proposals that result in lower monetary payoffs for the responder than the offer rejected. Indeed, such behavior can be part of a subgame-perfect equilibrium when the counter-proposal results in a sufficiently higher relative payoff for the responder. The model is less successful when fixed payments are added to the bargaining payoffs. For example, consider the far left point in Figure 1, which corresponds to the decision listed in the first column of Table 1. Since the pie remaining is zero after the first period this bargaining situation is a simple ultimatum game. The relative payoff model predicts that an offer of 25 cents will be accepted, since acceptance raises both the responder's absolute payoff from 25 to 50 and the relative payoff from 25/265 to 50/(265+225). Thus the proposer would have to offer at most 25 cents, or roughly 10 percent of the pie to ensure acceptance, but the average offer is close to 160 cents, roughly two-thirds of the pie.

Next, consider Matthew Rabin's (1993) notion of a "fairness equilibrium" that incorporates intentions into a Nash analysis. The basic idea is that people like to help those who are helping them, and to hurt those who are hurting them. In his model, players' utility functions therefore depend on both monetary payoffs and "kindness functions" that reflect whether own and others' actions were intended to help or hurt. Rabin's model technically applies only to normal-form games, but it is intuitively clear that in ultimatum games it can predict why responders turn down small offers if they are interpreted as hurtful acts by proposers.[5] The fixed payments in our experiment, however, are beyond the control of the proposers and responders and hence do not enter the kindness functions as defined in Rabin (1993).[6] Consequently, behavior is unaffected by these fixed endowments and, unlike the data pattern shown in Figure 1, initial offers are predicted to be positively correlated with the size of the pie remaining.

An alternative to "psychological" or "comparative" models is to allow for decision errors that are sensitive to their economic consequences. The logit rule, for example, specifies that when a player faces *m* options, the choice probabilities, p_i, are proportional to an exponential function of the associated expected payoffs π_i^e:

$$p_i = \frac{\exp(\pi_i^e/\mu)}{\sum_{j=1,..,m} \exp(\pi_j^e/\mu)}, \qquad i = 1,\dots,n, \tag{1}$$

where the sum in the denominator ensures that the probabilities sum to one. The "error parameter," μ, in (1) determines how sensitive choice probabilities are to payoff differences. As μ goes to infinity, the arguments of the exponential expressions go to zero, and the probabilities go to 1/*m*, regardless of expected payoffs. Thus a high μ represents noisy decision making that makes choices essentially random. In contrast, dividing expected payoffs by a low value of μ

---

5 Since intentions matter as much as monetary payoffs, the same model can explain why responders accept equally small offers when these offers are made by a computer (Blount, 1995).

6 Rabin's (1993) definition for the kindness functions would be as follows. When the proposer offers an amount *x* which she believes the responder will accept, the kindness measure is: $f_P = (\pi_R(x) - (\pi_R^h+\pi_R^l)/2)/(\pi_R^h-\pi_R^l)$, where $\pi_R(x)$ is the responder's payoff of accepting *x* and π_R^h, π_R^l are the highest and lowest possible payoffs for the responder when she accepts. Let the total pie to be divided be denoted by Π and the fixed payments for the proposer and receiver be denoted by ω_P and ω_R respectively. The proposer's kindness function is then: $f_P = ((\omega_R+x)-(\Pi/2+\omega_R))/((\Pi+\omega_R)-\omega_R) = (x - \Pi/2)/\Pi$, independent of the fixed payments.

means that payoff differences are blown up, making choice probabilities sensitive to payoff differences. Hence the "noisy best response" rule in (1) includes perfectly rational behavior and completely random behavior as limiting cases. The *logit equilibrium* (McKelvey and Palfrey, 1995) results by requiring that the belief probabilities that enter the expected payoff functions on the right side match the choice probabilities that result from the logit rule on the left side.[7]

The possibility of decision errors provides another explanation for behavior in bargaining games. Consider, for instance, a simple ultimatum game. Since the responder is almost indifferent between accepting a small offer or rejecting it, the logit rule in (1) stipulates that the probability of rejecting a small offer is close to a half. Small offers thus result in low expected payoffs for proposers who are better off offering more, and by (1), such larger offers are thus more likely to occur.

Notice, however, that the logit choice probabilities in (1) remain unchanged when a constant is added to expected payoffs of all options. As a result, the fixed payments used in the experiment have no effect in a logit equilibrium. In fact, the logit equilibrium prediction for the relationship between initial offers and the amount of pie remaining lies somewhere between the positively sloped 45-degree line predicted by the Nash equilibrium ($\mu = 0$) and the flat equal-division line resulting from purely random behavior ($\mu = \infty$). Logit decision errors alone cannot explain the downward sloping pattern in Figure 1.

To summarize, none of these approaches on their own can explain the negative correlation between initial offers and the amount of pie remaining that is caused by the additional fixed payments. In the next section we show that a modification of Bolton’s (1991) comparative model together with logit decision error provides a remarkably good description of the data.

4. A Model with Asymmetric Inequality Aversion and Logit Decision Error

Recall that in Bolton’s comparative model a player cares about inequity only when others are better off (in terms of monetary payoffs). Recent papers of Bolton and Ockenfels (1999) and Fehr and Schmidt (1999) amend this formulation by incorporating a more general taste for equity

---

[7] The logit equilibrium has been successfully applied to explain behavior in a variety of environments (e.g. McKelvey and Palfrey, 1995; Ochs, 1995; Anderson, Goeree, and Holt, 1998; Capra *et al*., 1999; Goeree and Holt, 1999).

into players' utility functions. For two-player games, Fehr and Schmidt model player *i*'s utility as:

$$U_i(\pi_i, \pi_j) = \pi_i - \alpha_i \max(\pi_j - \pi_i, 0) - \beta_i \max(\pi_i - \pi_j, 0), \quad i, j = 1, 2,\ i \neq j, \qquad (2)$$

where π_i is player *i*'s monetary payoff and $0 \le \beta_i < 1$.[8] The second term on the right side of (2) measures the utility loss from disadvantageous inequality ("envy"), while the third term measures the disutility from advantageous inequality ("guilt"). The presumption is that inequality aversion is asymmetric: $\alpha_i \ge \beta_i$, i.e. a player suffers more from disadvantageous inequality.[9]

To illustrate the difference between (2) and Bolton's (1991) comparative model, consider again the far left point of Figure 1, which corresponds to the decision in the first column of Table 1. Because of the difference in fixed payments, the proposer always ends up with a higher monetary payoff than the responder, so only the first and third terms of the right side of (2) play a role in this case. In particular, an offer of *x* provides the proposer with a favorable payoff advantage of $\$2.40 - 2x$, so the proposer's utility is: $U_P = \$2.65 + \$2.40 - x - \beta_P(\$2.40 - 2x)$. When $\beta_P > 1/2$, this utility is *increasing* in the offer amount, and the proposer would prefer to give away the whole pie; since this offer also makes the responder best off, it will be accepted. When $\beta_P < 1/2$, the proposer would prefer lower offers, but has to worry that the responder may turn down offers that are too small. The optimal offer will in general depend on the responder's "envy parameter," α_R, but is always less than half the pie.[10]

In the experiment, four subjects offered the entire pie and six offered less than half the pie. Most of the offers, however, were outside this predicted range (i.e. 13 out of 23 decisions). One way to explain this discrepancy is to introduce some (logit) decision error into the decision making process. Specifically, a noisy version of Fehr and Schmidt's model is obtained by

---

[8] The non-negativity constraint means that a player does not like to be better off than others. If $\beta_i \ge 1$ then player *i* would be willing to throw away money in order to reduce inequality.

[9] See Bolton and Ockenfels (1999) for a related model.

[10] To see this, note that by accepting a fifty-fifty offer, the responder's monetary payoff goes up while the inequality caused by the fixed payments is not further increased. It is straightforward to show that, as a function of α_R, the optimal offer is given by $\$2.40 * \alpha_R/(1+2\alpha_R)$, which is less than \$1.20 for $\alpha_R \ge 0$.

replacing the monetary expected payoffs in (1) by the utility functions given in (2). To end up with a parsimonious model we shall assume that all responders have the same utility and error parameters, as do all proposers. These parameters can be estimated from the data using maximum likelihood techniques. The estimates so obtained reveal that the α and μ parameters are not significantly different for proposers and responders, but that their β parameters are. Table 2 summarizes the estimation results when this restriction is implemented.

Table 2. Maximum-Likelihood Estimates

Variable	$\mu_P = \mu_R$	$\alpha_P = \alpha_R$	β_P	β_R
estimate (standard error)	.55 (.06)	.84 (.16)	.66 (.08)	.12 (.02)

To get a feel for how well this model fits the data, these estimates are used to calculate the distributions of offers and counter-offers using Fehr and Schmidt's utility function in (2) dressed up with logit error. Figure 2 shows the means of the resulting initial offer distributions, and Figure 3 shows the means of the counter-offer distributions, both as fractions of the $2.40 pie. The predictions of the three parameter model (dark lines) track the negatively sloped data average line in one case and the upward sloped data average line in the other (thin lines).

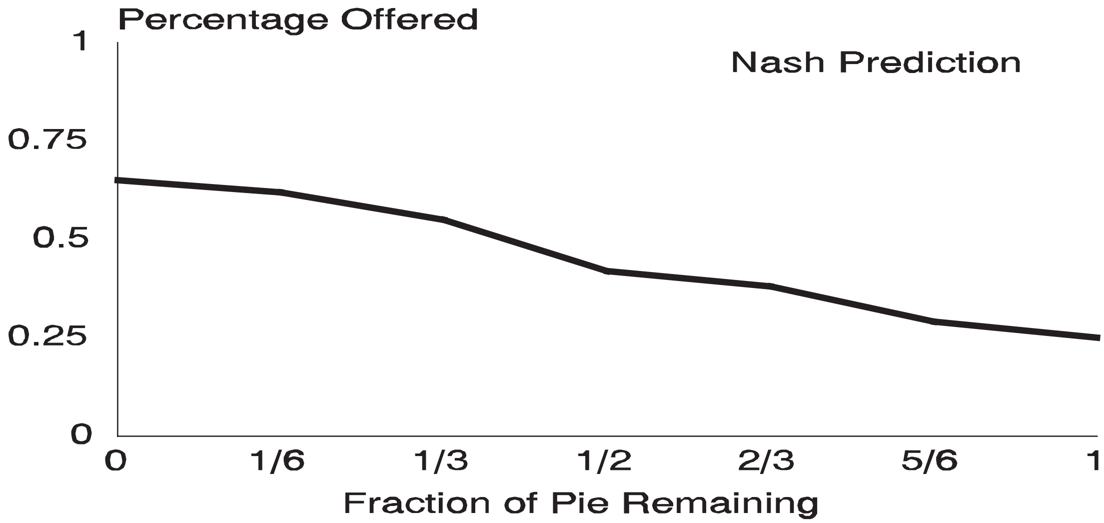

Figure 2. Initial Offers as a Fraction of the Pie:
Logit Prediction (dark line) and Data Averages (light line)

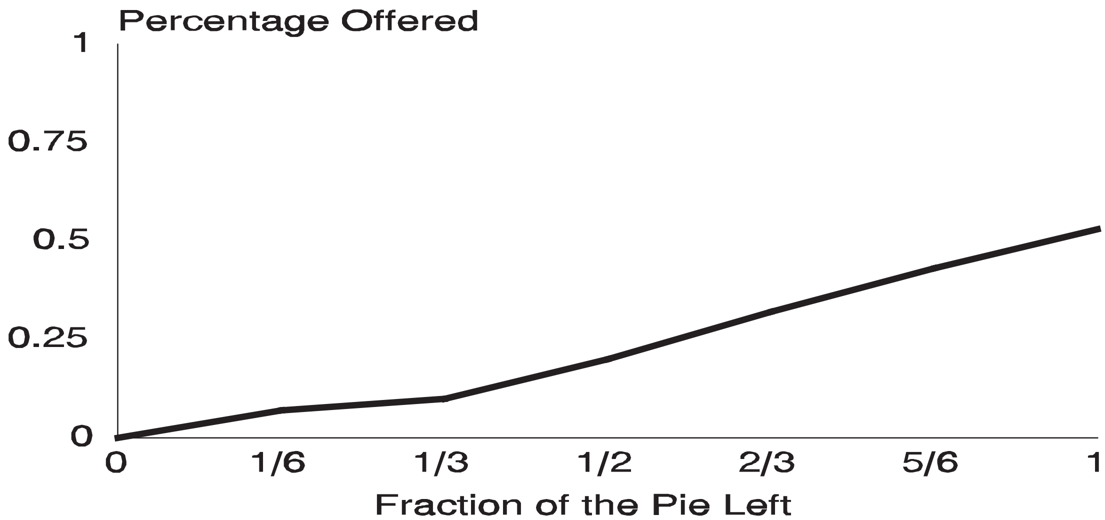

Figure 3. Responder's Counteroffers as a Fraction of the Initial Pie:
Logit Predictions (dark line) and Data Averages (light line)

Conclusion

Data from bargaining games have revealed an interesting mix of seemingly rational and irrational behavior. In fact, many experimental economists have considered the possibility that rejections in ultimatum games might be anomalous artifacts of experimental design that could be corrected with proper controls and anonymity. This hope was reinforced by the observation that initial offers in such games seemed to be approximately expected-payoff-maximizing choices conditional on rejection probabilities. This point of view has now been largely abandoned, since at least some anomalies persist even with high incentives, repetition, reductions in social distance, etc.

In a nutshell, most deviations from subgame perfect Nash predictions in bargaining experiments are in the direction of equal splits (Roth, 1995). This suggests two approaches: 1) keep the Nash equilibrium structure with backward induction rationality, and incorporate "Fehrness" and equity considerations into individuals' utility functions, or 2) keep selfish preferences and relax the Nash assumption of perfect payoff maximization, since noisy behavior typically biases behavior from extreme equilibrium demands towards equal division. Fortunately, these two approaches are not mutually exclusive and can be nested in a general model of asymmetric inequality aversion and probabilistic choice. This model is tested with an experiment in which asymmetric fixed money payments are specified so that equal division can only be reached by extreme demands for some treatments. Consequently, the equal division outcomes are arrayed along a line that is the *inverse* of the perfect positive relationship between remaining pie size and initial offers in a subgame perfect Nash equilibrium. This inverse relationship is observed in the data and is reproduced by the theoretical model, with the error and inequality aversion parameters estimated from the data with maximum likelihood methods. The model explains key aspects of the data that are inconsistent with models that rely only on noisy behavior or inequality aversion.

References

Anderson, Simon P., Jacob K. Goeree, and Charles A. Holt (1998) "A Theoretical Analysis of Altruism and Decision Error in Public Goods Games," *Journal of Public Economics*, *70*, 297-323.

Blount, Sally (1995) "When Social Outcomes Aren't Fair: The Effect of Causal Attributions on Preferences," *Organizational Behavior and Human Processes*, *63*, 131-144.

Bolton, Gary E. (1991) "A Comparative Model of Bargaining: Theory and Evidence," *American Economic Review*, *81* (December), 1096-1136.

Bolton, Gary E. and Axel Ockenfels (1999) "A Theory of Equity, Reciprocity, and Competition," forthcoming in the *American Economic Review*.

Capra, C. Monica, Jacob K. Goeree, Rosario Gomez, and Charles A. Holt (1999) "Anomalous Behavior in a Traveler's Dilemma?" *American Economic Review*, *89*(3), June, 678-690.

Davis, Douglas D. and Charles A. Holt (1993) *Experimental Economics*, Princeton, N.J.: Princeton University Press.

Fehr, Ernst and Klaus M. Schmidt (1999) "A Theory of Fairness, Competition, and Cooperation," forthcoming in the *Quarterly Journal of Economics*.

Goeree, Jacob K. and Charles A. Holt (1999) "Stochastic Game Theory: For Playing Games, Not Just For Doing Theory," forthcoming in the *Proceedings of the National Academy of Sciences*.

McKelvey, Richard D. and Thomas R. Palfrey (1995) "Quantal Response Equilibria for Normal Form Games," *Games and Economic Behavior*, *10*, 6-38.

Ochs, Jack and Alvin E. Roth (1989) "An Experimental Study of Sequential Bargaining," *American Economic Review*, *79*, June, 202-217.

Ochs, Jack (1994) "Games with Unique, Mixed Strategy Equilibria: An Experimental Study," *Games and Economic Behavior*, *10*, 202-217.

Rabin, Matthew (1993) "Incorporating Fairness into Game Theory and Economics," *American Economic Review*, *83*(5), December, 1281-1302.

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