

Bargaining and Reputation: An Experiment on Bargaining in the Presence of Behavioral Types*

Matthew S. Embrey Guillaume R. Fréchette
Maastricht University New York University

Steven F. Lehrer
Queen's University and NBER

August 22, 2012

Abstract

We conduct a series of laboratory experiments to understand what role commitment and reputation play in bargaining. The experiments implement the *Abreu and Gul* (2000) bargaining model that demonstrates how introducing behavioral types, which are obstinate in their demands, creates incentives for all players to build reputations for being hard bargainers. The data are qualitatively consistent with the theory, as subjects mimic induced types. Furthermore, we find evidence for the presence of complementary types, whose initial demands

*We wish to thank Jim Andreoni, Colin Camerer, John Kagel, Muriel Niederle, Tom Palfrey, Ennio Stacchetti, Leeat Yariv, as well as seminar participants at Caltech, Maastricht University, New York University, Royal Holloway, Universitat Autònoma de Barcelona, Università Bocconi, University of Michigan and University of Texas-Austin, and the ESA Asia-Pacific meetings, ESA North American meetings, the FSU Workshop on Experimental Game Theory, the Stony Brook Game Theory Festival, and the SITE Conference on Experimental Economics for helpful comments and suggestions that improved the paper. We thank Raj Advani and Samuel Mati for software development. Fréchette and Lehrer respectively wish to thank the NSF (grants SES-0519045, SES-0721111, and SES-0924780), CV Starr Center, Center for Experimental Social Science (CESS) and SSHRC for research support. We are responsible for all errors.

acquiesce to induced behavioral demands. However, there are quantitative deviations from the theory: subjects make aggressive demands too often and participate in longer conflicts before reaching agreements. Overall, the results suggest that the *Abreu and Gul* (2000) model can be used to gain insights to bargaining behavior, particularly in environments where the process underlying obstinate play is well established.

1 Introduction

Since bargaining is the process through which many agreements are reached within the economy, it has been an active area of research in political science and industrial relations, and among theoretical, empirical and experimental economists. A common feature of many bargaining situations is that bargaining parties attempt to increase their share of the surplus by claiming they are committed to a particular bargaining position.¹ Such commitment tactics have been recognized as not only an important factor in determining bargaining outcomes, but also as a source of conflict. Consequently, understanding the role of commitment and reputation in the bargaining process not only provides a potential explanation for the differences in allocations received by agents, but may also account for the inefficiencies that result from delay and disagreement.²

Such posturing has been incorporated into a number of bargaining models, including the classical non-cooperative theory of *Rubinstein* (1982). This literature has culminated in the model of bargaining and reputation by *Abreu and Gul* (2000).³ In their stylized game, players bargain in two stages. In the first stage, the two players choose their bargaining positions, that is the share

¹The classic essay by *Schelling* (1960) developed the literature on strategic commitment, documenting many examples of such behavior.

²In the current context, commitment is understood to mean that an agent can only agree to a predetermined outcome or set of outcomes. Commitment does not result from the timing protocol and differs from Stackelberg leadership. Reputation in this paper refers to incomplete information over whether a player is committed to some predetermined outcome or set of outcomes. This arises since there is uncertainty about the types of both players, which differs from models (and much of the previous experimental literature) in which only one actor can build a reputation.

³The role of strategic commitment was first formulated into a non-cooperative model by *Crawford* (1982). *Chatterjee and Samuelson* (1987, 1988) were among the first to incorporate two-sided reputation.

of the pie they demand. If the two positions are compatible with each other, the game ends. If not, no further offers are allowed, and the players enter the second stage: a continuous-time concession game. The second stage ends when one player concedes to the demand of the other. Reputation enters through the possibility that a player is behavioral. A behavioral player will never concede to any offer that does not give them at least the amount that they demanded in the first stage.⁴

In this paper, we implement this stylized bargaining model in the laboratory to investigate the role of commitment and reputation in bargaining environments. Although the two-stage format may appear very abstract, the design has the advantage of eliminating a potential confound on bargaining power, namely proposer power, that arises in other protocols. This abstraction allows the experiment to focus on the role played by reputation in generating bargaining power, while providing the structure needed for testable predictions on players' strategies. The symmetric setting also provides a tension between following the simple 50-50 bargaining norm, which is fair and efficient, and the demands of reputation, which can lead to unequal offers and delay in equilibrium. Furthermore, *Abreu and Gul* (2000) demonstrate with their convergence result that bargaining power based on reputation, as in the stylized game, is the underlying strategic interaction for a general class of bargaining environments, as one abstracts from the particulars of the bargaining protocol.⁵

Viewing the strategic interaction in the stylized two-stage format offers a fresh perspective on the bargaining problem, one that has not previously been brought to the laboratory. Indeed, the experiments reveal new insights into subject behavior: first, there is more conflict than in previous bargaining studies; second, subjects fight for their demands more than predicted, and third, while there is clear evidence that subjects recognize the role of reputation in this environment, they tend to be more aggressive in their demands than predicted. In all cases, these observations result in less efficiency and less equality, unlike prior bargaining studies in which these are the two pri-

⁴The r-insistent types of *Myerson* (1991) are an example of the earlier use of such types.

⁵The convergence result connects the equilibrium outcomes of the stylized game with that of the limit of discrete-time bargaining games, as the time between bargaining rounds goes to zero. This result allows for a general class of bargaining protocols, including alternating offers as a special case. Consequently, as the time between bargaining rounds becomes smaller, the dominant factor for outcomes is reputation, as captured by the stylized game, rather than some feature of the protocol.

mary forces identified as generating deviations from equilibrium predictions.

While the theoretical considerations outlined above motivated our interest in testing the stylized game in the laboratory, it does not motivate the addition of irrationality in the model. The formulation of the behavioral types permits a tractable model that highlights the role of strategic posturing and delivers delay as an equilibrium outcome.⁶ There are several ways in which to view these types. First, these types are genuinely boundedly rational, in the sense that they are obstinate, or follow a rule of thumb or bargaining convention that has evolved outside the current model. Indeed there is experimental support for such an interpretation. An example of such behavior is the observation of disadvantageous counter-offers identified by *Ochs and Roth* (1989). Following a rejection in alternating-offer bargaining games with a shrinking pie, the counter-offer often gives less to the subject that rejected the initial offer. This is despite the fact that the counter-offer is made by that very same subject. The types included in *Abreu and Gul* (2000) would also make disadvantageous counter-offers in such environments. A second interpretation of these types is that they result from actions, taken before the bargaining process has begun, that could potentially commit a player to being unable to accept less than a certain outcome.⁷ In both cases, either bounded rationality or commitment actions, there is the possibility that either player is irrevocably bound to a demand. The current study will not attempt to determine the nature of behavioral types in the population, but rather, will focus on the extent to which behavior accords to the theory given their presence.

An empirical investigation of the model is of interest, not only for its novel perspective on the bargaining problem, but also for the impact it has had on the economics literature. In addition to studies that have proposed theoretical extensions - for example *Wolitzky* (forthcoming) and *Atakan and Ekmekci* (2010) - the results have been applied to a diverse range of topics: insights from the model have extended the literatures on negotiations (*Damiano et al.* 2012), political competition over entitlements (*Baron* 2003), banking (*Povel* 2005), auction markets (*Kwiek* 2011), and issues arising in imperfectly competitive search markets (*Özyurt* 2011). The laboratory is particularly suited

⁶See *Crawford* (1982) for an exposition on the need for both uncertainty and irrevocability in a bargaining model with commitment and delay.

⁷See *Muthoo* (1999) or *Kambe* (1999) for models that demonstrate that incomplete information over whether such actions have been taken and their availability, leads to a similar environment to that in *Abreu and Gul* (2000).

for our investigation given a number of challenges that would arise with field data: first, the predictions of the model rely on knowledge of the sample distributions of both rational and behavioral types.⁸ Second, to distinguish between alternative motives for conflict and delay, it is necessary to identify the potential types of the agents and their payoff functions. Finally, it is difficult to imagine a natural experiment where everything but the possibility for building reputation remains constant. In the laboratory it is possible to create a controlled environment where the number and the demands of behavioral types are the sole difference between sessions, thereby providing a direct test of the theory.

The paper presents analyses from five treatments. The treatments vary the demands that can be made: either any feasible demand (the unrestricted design) or only a predetermined subset (the restricted design). The other variation is the introduction of computer players that follow a pre-specified strategy (referred to as induced behavioral types).⁹ The unrestricted design aims to determine if the subjects recognize the role of posturing in bargaining. This will be the first, and in our view the key, comparative static: do the subjects mimic the induced behavioral types when they should (and do not when they should not). Although this is a fairly basic test of the theory, laboratory data routinely reject basic predictions of standard models and, in this case, the prediction runs counter to other decision heuristics. In the restricted design, subject demands are restricted to those made by induced types, to ensure that the set of behavioral types in the treatment is common knowledge for all subjects. These modifications allow us to conduct more precise tests of the finer details of the model, in particular with regard to the length of delay and the pattern of concession in the second stage.

Our results provide clear qualitative support that subjects recognize the role of strategic posturing by mimicking induced behavioral types: across treatments, subject demands for induced types that are predicted to be mimicked increase, while there is no significant change in demands of induced types not predicted to change. In addition, the experiment reveals

⁸Even with demand data, rather than outcome data, these sample distributions cannot be simply estimated since there are too many unrestricted parameters.

⁹We induce certain behavioral types to provide experimental control over the predictions of the model. In *Roth and Schoumaker* (1983) subjects also bargain with computers, but they are not aware of it. Other experiments involving computer players, but in different strategic settings, include *Andreoni and Miller* (1993) and *Grosskopf and Sarin* (forthcoming), among others.

behavior that, although not predicted by this model, is consistent with an evolutionary extension of the model. Namely, we find evidence for the presence of complementary types (*Abreu and Sethi 2003*), whose initial demand instantaneously acquiesce to induced obstinate demands. However, there are quantitative differences between the observed behavior and several of the finer details of the model, in particular with the length of delay and the pattern of concession in the second stage play. We conjecture that these deviations are due, in part, to some uncertainty over the set of behavioral types in the subject population. The behavior in the restricted design is mostly consistent with this hypothesis, although some disparities do persist: subjects appear to be disproportionately mimicking more aggressive types in the first stage, and delays in the second stage remain longer than predicted. A minor modification to the assumed behavior of obstinate types would result in predictions with more demanding first-stage announcements and greater delay, thus accommodating some of the identified disparities.

The paper is organized as follows. In the next section we provide a brief review of the related experimental literature. Section 3 outlines the theory behind a symmetric version of the bargaining and reputation model implemented in the laboratory, and highlights the relevant predictions from this model. Section 4 introduces the experimental design. Section 5 presents the results, a discussion of which is given in section 6. A final section draws the main conclusions.

2 Related Experimental Literature

The results presented in this paper relate to experimental studies of bargaining, reputation and war of attrition models.¹⁰ With regard to bargaining, a large literature has documented observations of norm-driven behavior by subjects. For instance in bilateral bargaining, agents frequently agree on rather egalitarian outcomes in situations where the standard model, with purely selfish preferences, predicts rather unequal outcomes. In his summary of the literature, *Roth (1995)* also notes a number of observable regularities

¹⁰To the best of our knowledge, the war of attrition game has received very little attention in the experimental literature, and has not been implemented in the manner in which it is here. However, there is a formulation that can be represented as an all-pay auction (*Bulow and Klemperer 1999*), which has been the subject of a number of experimental studies that are surveyed in *Hörisch and Kirchkamp (2010)*.

in bargaining regarding the establishment of focal points, what determines credible bargaining positions, and the role of subjects' expectations of one another as they develop during sessions.¹¹

These experimental findings led to the development of models incorporating other-regarding preferences, which initially focussed on concerns for inequality aversion (*Fehr and Schmidt* 1999, *Bolton and Ockenfels* 2000).¹² This literature points in a different direction from the research of Roth and his co-authors and suggests a different cause for conflicts in bargaining: fairness preferences or feelings of spite, rather than strategic posturing.¹³ Finally, there is a related experimental literature on bargaining impasses in negotiations, which postulates that such impasses can result from subjects conflating fairness and best interest (see for instance *Babcock and Loewenstein* 1997).

The experimental literature on reputation building has focussed mainly on testing models of one-sided incomplete information in a repeated game setting. A series of experiments have tested the sequential equilibrium predictions of variants of the borrower-lender (trust) game, which was first implemented in the laboratory by *Camerer and Weigelt* (1988).¹⁴ While there are several differences in theoretical structure from these earlier experiments, the most important is that the model implemented in the experiments presented here has two-sided incomplete information. This means that the incentive to build reputation is symmetric, as is the need to incorporate the other player's incentive to build reputation into one's own strategy. An earlier experiment that does incorporate two-sided incomplete information is that by *McKelvey and Palfrey* (1995). In their study of a repeated hold-out game, a move of nature at the beginning of a match determines the types of the two subjects, which in turn determines which stage game they are playing. An important

¹¹Studies by *Roth et al.* (1981), *Roth and Murnighan* (1982), *Roth and Schoumaker* (1983) have shown that, by increasing the number of norm focal points (besides 50/50), experimental outcomes become less concentrated at 50/50 and increase the probability that the bargainers will fail to reach an agreement.

¹²This continues to be an active area of research, with the role of intentions (*Charness and Rabin* 2002) receiving increasing attention. See *Cooper and Kagel* (forthcoming) for a review of this literature.

¹³For example, *Falk et al.* (2003) examine reputation in ultimatum games defined as the recent history of acceptance/rejection of past offers. They conclude that, while it changes the outcome, the effect of reputation is to reduce the proposers' uncertainty about the responders' acceptance thresholds.

¹⁴See *Neral and Ochs* (1992), *Brandts and Figueras* (2003) and *Grosskopf and Sarin* (forthcoming). *Jung et al.* (1994) use a similar design but test the chain store game.

behavioral implication of their equilibrium is for subjects to infer from the history of play which stage game they are playing, a feature that does not have a direct analog in the *Abreu and Gul* (2000) model. Overall, this strand of the literature observes a broad notion of reputation building by subjects in their respective environments.¹⁵ However, the finer details of sequential equilibrium predictions can often fail to be borne out in the data.

In addition, it is unclear whether the evidence of reputation building in these other environments will translate to a general bargaining situation. A broad interpretation of the results from previous studies of one-sided reputation would suggest that subjects would be inclined to mimic the first stage demands of behavioral types, which corresponds to reputation building in the current context. However, reputation building in our experiments will not always align itself with adherence to some common bargaining norm (such as the 50-50 split), yet following such a norm will always be an available, and more efficient, option for subjects. Consequently, although experiments of one-sided reputation models have found some support, it is not clear that this will extend to a model such as the one considered here where the equilibrium predictions run counter to both the norms of fairness and efficiency.

3 Summary of the Theory

This section provides an overview of the relevant theoretical results for a symmetric version of the stylized bargaining and reputation model of *Abreu and Gul* (2000).¹⁶ Two agents bargain over a pie of size one in two stages. In the first stage (at time 0) each player simultaneously announces a posture α^i (i.e. the fraction of the pie they would like). If the two suggestions are compatible (i.e. $\alpha^1 + \alpha^2 \leq 1$) then the game ends immediately.¹⁷ If the two suggestions are incompatible, the game proceeds to stage two, where a continuous-time concession game with an infinite horizon starts. That is, for each point in time, $t \in [0, \infty)$, both players choose to accept (i.e. concede)

¹⁵An exception is *Brandts and Figueras* (2003) who find that the predictions of the sequential equilibrium do not organise their results well.

¹⁶What is contained here is implicit in *Abreu and Gul* (2000). Some results for equilibrium announcements are drawn from *Abreu and Sethi* (2003), where the authors use the symmetric version of the model.

¹⁷If the announcements sum to strictly less than one, a sharing rule is used for the remainder. While other sharing rules can be accommodated, in the experiments we divide the remainder equally.

or hold out. If player i concedes, she receive $1 - \alpha^j$, while if j concedes, player i receives α^i . Preferences of agents are risk neutral, with a common discount factor r . Thus, if an agreement is reached at time t in which an agent receives a share x , then their payoff is $e^{-rt}x$.

In addition, there is some probability that a player may face a behavioral type who is obstinate in their demands. Define $C := \{\alpha_1, \dots, \alpha_K\}$ as the set of behavioral types, with $\alpha_1 + \alpha_K \geq 1$. An α_k -type always demands α_k and only accepts an offer that gives them at least α_k . The probability that a player is an α_k -type is z_k , for $k = 1, \dots, K$. The probability that a player is rational is denoted by $z_0 = 1 - \sum_{k=1}^K z_k$. Last, a behavioral type is defined as *aggressive* if it is incompatible with all other behavioral types in C including itself.

3.1 Equilibrium Behavior

A key property of the equilibrium is that a rational player would only choose a demand that mimics some behavioral type (i.e. $\alpha^i \in C$ for $i = 1, 2$). To do otherwise would instantly reveal the rationality of the actor and result in their opponent appropriating any gains from trade. Consequently, players can be identified by the element of C that they announce in the first stage, $\alpha_k, \alpha_l \in C$. In a symmetric equilibrium, define μ_k to be the probability that a rational player announces demand α_k to mimic the behavior of α_k . Given this symmetric equilibrium, the probability that a player is irrational given an announcement α_k is given by

$$\bar{\pi}_k = \frac{z_k}{z_k + z_0 \mu_k} \quad (1)$$

In equilibrium, with a general set C , rational players will employ a mixed strategy over announcing different types in first stage. If the set C contains a type, α , such that $\alpha \leq \frac{1}{2}$, there is a possibility that this type will not be replicated in the equilibrium mixed strategy. However, if a behavioral type is replicated, then all more aggressive types are also replicated. As such, the support of the equilibrium mixing strategy, μ , will be of the form $\{\alpha_R, \dots, \alpha_K\}$, where $1 \leq R < K$. Ensuring that rational players are indifferent between announcing any α_k (for $k = R, \dots, K$), along with the μ being a probability measure (and therefore summing to one), yields the $(K - R + 2)$ equations needed to solve for μ and the expected payoff for rational players.¹⁸

¹⁸This is a system of non-linear equations that only has a numerical solution, except for

Suppose a rational player announced α_k and faces an opponent who has announced α_l , where $\alpha_k + \alpha_l > 1$, causing the players to move on to the concession stage.¹⁹ The unique equilibrium play in the incomplete information war of attrition game is given by a mixed strategy over the time of concession. The α_k player concedes with constant hazard rate, λ_{kl} , given by²⁰

$$\lambda_{kl} = \frac{r(1 - \alpha_k)}{\alpha_k + \alpha_l - 1} \quad (2)$$

over the interval $[0, T_0]$, where $T_0 = \min(T_{kl}, T_{lk})$ and $T_{kl} = \frac{-\ln(\bar{\pi}_k)}{\lambda_{kl}}$ and $T_{lk} = \frac{-\ln(\bar{\pi}_l)}{\lambda_{lk}}$. Thus, equilibrium is generally characterized by inefficient delay. Concession by the α_k rational player is governed by the distribution function $\frac{\hat{F}^{kl}}{1 - \bar{\pi}_k}$, where

$$\hat{F}^{kl}(t) = \begin{cases} 1 - c_{kl}e^{-\lambda_{kl}t}, & \text{for } t \in [0, T_0] \\ 1 - \bar{\pi}_k, & \text{for } t > T_0 \end{cases} \quad (3)$$

and $c_{kl} = \bar{\pi}_k e^{\lambda_{kl}T_0}$ and $(1 - c_{kl})(1 - c_{lk}) = 0$. Note that the distribution function is expressed in terms of \hat{F}^{kl} for notational convenience: a rational player who announced α_l , when their opponent announced α_k , faces a “mixed” strategy over the time of concession given by \hat{F}^{kl} (i.e. unconditional on the α_k player being rational).

The value of T_{kl} is a measure of the α_k rational player’s “strategic” weakness when facing an α_l player: if $T_{kl} > T_{lk}$, then the α_k rational player will have to concede at time $t = 0$ with strictly positive probability (mass), given by $q_{kl} := (1 - c_{kl})$. Such concession is referred to as *initial* concession. Concession resulting from the continuous part of the distribution function is referred to as *interior* concession. Finally, revisiting the first stage, if a type is aggressive then it is always replicated in equilibrium and never conceded to initially by another player.

trivial cases. Section 1.7 of the online appendix contains details of the numerical strategy used to solve this system of equations for a general set of behavioral types.

¹⁹Note that the player who announced α_l could be either an α_l -type or a rational player who has mimicked the α_l -type.

²⁰So long as it remains possible that their opponent is a rational-type, a rational player who announced α_k is indifferent between conceding and not conceding at a time t if $r(1 - \alpha_l) = [\alpha_k - (1 - \alpha_l)]\lambda_{lk}$, where λ_{lk} is the hazard rate for concession by the opponent unconditional on knowing whether the opponent is rational or not. Equation 2 ensures this indifference holds.

3.2 Key Equilibrium Predictions

Although the exact nature of equilibrium play by rational players - and the consequent equilibrium outcomes - is dependent in a non-linear manner on the set of behavioral types and the distribution over these types, there are a number of features of equilibrium behavior that hold irrespective of such parameters.²¹ With regard to first-stage announcements made by rational players, the following can be said:

1. Rational players will only make announcements that mimic some behavioral type.
2. If the announcement of a behavioral type is mimicked in equilibrium, then the announcements of all more demanding behavioral types are also mimicked.

With regard to second-stage behavior by rational players in the subgame following announcements α_L and α_H , where $\alpha_H + \alpha_L > 1$ and $\alpha_L \leq \alpha_H$:²²

3. The upper bound on the average (mean) delay, given that agreement is eventually reached, is $\frac{1}{\lambda_{LH} + \lambda_{HL}}$.
4. If $\alpha_L = \alpha_H$ or if both α_L and α_H are aggressive, there is no initial concession by either player.
5. If $\alpha_L < \alpha_H$, then the rational player who initially announced α_L would not concede initially. Whether a rational α_H announcer would initially concede depends on the full solution of the first stage equilibrium mixing strategy. However, this will generally be the case if α_H is not aggressive. Furthermore, $\alpha_L \leq \alpha_H$ implies that $\lambda_{LH} \geq \lambda_{HL}$. Consequently, unconditional on knowing whether or not the players are rational, the more demanding α_H announcer concedes at a slower rate than the α_L announcer for interior concession.

²¹See sections 1.4, 1.5 and 1.6 of the online appendix for further details on the theoretical predictions for announcement behavior, delay and concession behavior.

²²Throughout the paper, the term *subgame* is used to refer to the information set reached following incompatible first-stage announcements. This notation is an abuse of terminology, since the incomplete information means these are not proper subgames, but is used for expositional ease.

4 Experimental Design

In order to implement the stylized game in the lab, there are three features of the model that need to be induced. First, computer players are used to gain some control over the set of behavioral types and the distribution over this set. The computer players are programmed to follow a fixed α -rule and to never concede in the second-stage. To make this information common knowledge among the subjects, the rule and the probability of being matched with a computer player is included explicitly in the instructions.²³ Second, following the tradition of *Roth and Malouf* (1979), risk neutral preferences are induced using the lottery method. During the experiment, subjects bargain for probability points, as opposed to monetary amounts; a binary lottery is conducted at the end of the session, where the probability of winning is given by the number of probability points won during the session as a fraction of the total possible number of points.²⁴ Third, to induce the common discount rate, the pie is shrunk continuously over time according to the rate $r = 0.01$ per second.

The experiments were conducted at the CESS laboratory at NYU using undergraduate students from all majors, who were recruited through e-mails. All treatments were implemented using a between subjects design. Instructions were read to students aloud and they interacted solely through computer terminals.²⁵ The unrestricted design was as follows: Subjects were randomly matched in pairs, then simultaneously placed a demand (between 0 and 30). If the demands they made were compatible with the demand of the person they were matched with (that is, summed to 30 or less), then the round ended and each subject earned their demand plus half of any remainder. If the demands were incompatible, then that match moved on to a second stage. This stage proceeded in continuous time, with subjects having the option to choose to concede. If either they or the person they

²³An alternative would have been to use human subjects who have a different payoff structure, namely they receive α points if and only if they get α in the bargaining and 0 otherwise. Since it is not clear what subjects would actually do in this situation, computer players were used instead to maintain control over the environment. Furthermore, this would generate concerns regarding what other subjects believe about how these subjects would choose to behave.

²⁴This method induces risk neutral preferences over probability points, regardless of the subject's attitudes towards risk in monetary payoffs.

²⁵See sections 2.1 and 2.2 of the online appendix for sample instructions and screen shots.

were matched with conceded then the round ended. The subject who conceded earned 30 points minus the demand of the player with whom they were matched with, all multiplied by the discount factor for the amount of time taken to reach an agreement ($e^{-0.01t}$ where t is the time in seconds at which one of the two subjects conceded). The subject that did not concede earned their demand multiplied by the discount factor. Throughout the second stage, subjects were shown a 2×2 matrix. This displayed, in real-time, the discounted payoffs to each player, both if they conceded at that moment or if the other player conceded at that moment. During each session, subjects were randomly rematched 15 times to provide subjects an opportunity to gain experience with the game. Sessions lasted approximately one and a half hours.

Treatment		Session			Total
		1	2	3	
<i>C0</i>	Number of Subjects	12	12	10	34
	Number of Computers	0	0	0	
	Computer Types	.	.	.	
	Mean Earnings (\$)	25.0	23.3	23.0	23.8
<i>U1</i>	Number of Subjects	14	14	14	42
	Number of Computers	2	2	2	
	Computer Types	20	20	20	
	Mean Earnings (\$)	20.7	23.6	22.1	22.1
<i>U2</i>	Number of Subjects	14	14	14	42
	Number of Computers	2	2	2	
	Computer Types	12, 20	12, 20	12, 20	
	Mean Earnings (\$)	22.1	25.0	27.9	25.0
<i>R3</i>	Number of Subjects	13	13	13	39
	Number of Computers	3	3	3	
	Computer Types	15, 18, 20	15, 18, 20	15, 18, 20	
	Mean Earnings (\$)	24.2	24.2	19.6	22.7
<i>R4</i>	Number of Subjects	12	12	12	36
	Number of Computers	4	4	4	
	Computer Types	8, 15, 18, 20	8, 15, 18, 20	8, 15, 18, 20	
	Mean Earnings (\$)	25.0	15.0	15.0	18.3

Table 1: Summary of All Sessions

Three treatments using the unrestricted design were conducted. These varied the presence and demands of computer players in a session. As indicated in table 1, the control treatment, referred to as *C0*, did not contain any computer players, while the second and third treatments, referred to as *U1*

and $U2$ respectively, included two computer players in each session. In the former both computer players were programmed with a demand of 20; in the latter one computer player was programmed with a demand of 12, the other 20. Table 1 provides a summary of the sessions. Under the assumption that the only behavioral types are the computer players introduced in $U1$ and $U2$, the model would predict the following: while there would be no precise prediction for the control,²⁶ in $U1$ there would be only announcements of 20, and in $U2$ there would be only announcements of 12 and 20.

However, given the evidence in bargaining experiments for norm driven behavior, it is reasonable to postulate that not all subjects who enter the laboratory correspond to a rational-type player of the model.²⁷ As discussed in Section 3, the exact predictions of the model are dependent on the set of behavioral types and their distribution. Consequently, the set of potential behavioral types (and their distribution) amongst the population of subjects must be considered, along with any induced types. However, for many assumptions over this set of types, moving from the control to either $U1$ or $U2$ would imply an increase in the number 20 demands, while the demand of the 12 computer would not be mimicked by rational subjects.

As an illustration, table 2 gives the predictions of the model for several example scenarios besides the case already outlined. The first scenario considers the possibility that on average one out of 13 subjects brought into the laboratory acts like a 15-type.²⁸ That is, they always demand the 50 – 50 split and would never accept less. As the last column of the table shows, while the probability of observing an announcement of 12 by a subject is not predicted to change across the treatments, the basic comparative static for announcements of 20 is the same. Namely, the treatment manipulation between the control and either $U1$ or $U2$ should result in an increase in

²⁶When there are no behavioral types in the model, i.e. in the complete information case, there exists a continuum of equilibria. Therefore, the outcome of this control treatment must be interpreted in conjunction with the other two treatments, where there are equilibrium predictions irrespective of whether the computer players are the only behavioral types.

²⁷This is somewhat in the spirit of the *homemade beliefs* of *Camerer and Weigelt* (1988), although in the current context there is two-sided asymmetric information.

²⁸If one of fourteen subjects, in either $U1$ or $U2$, were a 15 type, then the probability of one of the other subjects being matched with this player would be $1/15$. This is because there are also two computer players added to the matching pool. Without the two computer players, this probability would be $1/13$.

demands of 20.²⁹

Treatment		z			μ			Prob. Observing [§]		
		12	15	20	12	15	20	12	15	20
<i>Scenario A</i> ¹										
<i>C0</i>	Assumed	0	7.7	0	0	100	0	0	100	0
<i>U1</i>	Predicted	0	6.7	13.3	0	53.4	46.6	0	57.	43.
<i>U2</i>	Predicted	6.7	6.7	6.7	0	69.	31.	0	71.4	28.6
<i>Scenario B</i> ²										
<i>C0</i>	Estimated	1.2	33.6	6.	0	0.8	19.7	1.2	33.6	7.5
<i>U1</i>	Predicted	1.1	29.1	18.5	0	1.4	43.5	1.2	33.7	9.3
<i>U2</i>	Predicted	7.7	29.1	11.9	0	0	36.6	1.2	33.6	8.8
<i>Scenario C</i> ³										
<i>C0</i>	Estimated	1.2	6.9	2.1	0	43.9	8.8	1.2	34.	7.5
<i>U1</i>	Predicted	1.1	6.	15.1	0	24.5	41.	1.2	22.	27.3
<i>U2</i>	Predicted	7.7	6.	8.5	0	31.3	30.9	1.2	26.2	21.1

Table 2: Example Rationality Scenarios

[§] Probability of observing an announcement, excluding those made by computer players. All values in the table are probabilities represented as percentages.

¹ z_{15} assumed to be such that, in *U1* and *U2*, 1 out of the 15 other players is expected to be a 15-type (the 50 – 50 norm).

² z estimated using announcement data, assuming $z_0 = \frac{1}{13}$.

³ z estimated using announcement data, assuming $z_0 = \frac{8}{13}$.

Clearly it is unrealistic to control ex-ante what types might be possible, or considered possible, in the subject population. However, under the assumptions of the model, announcement data from the control session can be used to estimate the set of behavioral types and its distribution, identified up to a given value for the probability of being matched to a rational player.³⁰ Using these estimates, one can predict the implied probability of demands of 12 and 20 in *U1* and *U2*, and confirm the comparative static empirically.

²⁹This example is *indicative* of the more general scenario where 20 is *not* an element of the behavioral types in the population. If z_0 is not too small or if 20 is the most demanding type, then 20 announcements should be observed in both *U1* and *U2*, but not in the control.

³⁰A second estimation strategy, which used both first and second stage data to estimate the parameters, was also considered. However, the excessive delays in the control treatment, with respect to the theoretical predictions (see section 5.2), meant that using second-stage data to estimate z resulted in unrealistically low estimates for z_{15} . Estimating z using just the announcement data does not suffer this drawback, but cannot identify the parameter z_0 . See section 2.4 of the online appendix for further details of the estimation strategies used and the resulting predictions.

The results of this analysis are reported in table 2 as scenarios B and C respectively, where the former assumes a minimal probability of being matched to a rational subject and the latter a much larger value.³¹ As with the previous scenario, the basic comparative static prediction is maintained: subjects should mimic the 20 type more often in the computer treatments, while the 12 type should never be mimicked. Consequently, data consistent with this comparative static is taken as evidence for subjects reacting as predicted by the model.

In order to exert more experimental control over the set of possible behavioral types, a modified design, referred to as the restricted design, was also used. These experiments were the same as under the unrestricted design, with two exceptions:

- Subjects were only permitted to announce a demand from a restricted set of possible announcements. The finite set from which announcements could be made was included in the instructions to ensure the set was common knowledge amongst all participants.
- For each permitted announcement, a single computer player, for whom that demand was their strategic posture, was included in the session. Thus for each demand a strictly positive probability of being matched with a behavioral player of this type was induced. From an experimental design perspective, this addition has the advantage of not creating any asymmetry between the behavioral types, as would be the case if computers player only played a strict subset of the announcements in the restricted set.

Two treatments using this modified design were conducted. In the first, referred to as *R3*, only announcements from the set $\{15, 18, 20\}$ were allowed. The second, *R4*, only allowed announcements from the set $\{8, 15, 18, 20\}$. The characteristics of each treatment are also summarized in table 1.

Demands of 15 and 20 were included in both treatments to capture key aspects of the unrestricted design. These demands were the most common choice in each treatment. Furthermore, the demand of 15, being the equal

³¹Scenarios B and C report the estimation of z using just announcement data, assuming a value of $z_0 = \frac{1}{13}$ and $\frac{8}{13}$ respectively. This estimation strategy was also conducted with z_0 taking values in $\{\frac{1}{13}, \dots, \frac{8}{13}\}$. The direction of the comparative static does not change with the various assumptions on z_0 . See section 2.4 of the online appendix for further details.

split, is ex-ante the most natural candidate for a potential behavioral type in the subject population. The 20 demand replicates the induced behavioral type subjects were expected to mimic in the $U1$ and $U2$ treatments. The demand of 18 is included to test further the announcement strategies of subjects. In the $R3$ treatment, where all demands are aggressive, an observation of significantly more 20 announcements than 18 announcements can only be rationalized by the model if there are significantly more 20 types than 18 types in the subject population.³²

In the $R4$ treatment, a demand compatible with all other types, including the most aggressive type 20, was included. In equilibrium, the model predicts that such demands of 8 should *not* be made by rational types.³³ Consequently, the inclusion of 8 permits some measure of the magnitude of a potential demand induced effect resulting from including a computer type. Furthermore, the presence of a demand for strictly less than half the pie has implications for second stage concession behavior. In particular, rational players making aggressive demands should concede initially with strictly positive probability to any announcement that is strictly less than their demand. The prediction in $R3$ differs: there should be *no* initial concession in equilibrium.

5 Results

Throughout this section, the analysis will focus on data from the last 10 periods involving non-computer players.³⁴ Furthermore, all statistical tests will use standard errors clustered at the session level, which allow for arbitrary correlation between observations from the same session. Table 3 summarizes outcomes from the bargaining process in the five treatments. On average the difference between subject earnings across treatments is small, with matches

³²This result is a consequence of there being no initial concession by rational subjects when all behavioral types are aggressive. See section 1.3 of the online appendix for further details.

³³This strategy is strictly dominated by making the most aggressive announcement and, therefore, cannot be a member of the support of the equilibrium mixing strategy.

³⁴The main text focuses on the last 10 periods, by dropping the first 5 periods of data, as there is often an initial period of learning by subjects. The overall conclusions of the results are not altered by including the first five periods. See section 3 of the online appendix for a replication of the analyses contained in this section with all 15 periods of the data, as well as a detailed description of all the reported statistical tests.

in the unrestricted design wasting more of the pie, particularly treatments with induced types. While for all treatments the majority of matches enter a second stage, for the *U2* treatment and both restricted design treatments this rate is larger, at over 75%. To evaluate the performance of the model across the treatments, subject behavior is examined in three specific areas: first-stage announcements, second-stage delay and second-stage concession.

	Treatment				
	<i>C0</i>	<i>U1</i>	<i>U2</i>	<i>R3</i>	<i>R4</i>
Mean points per period	13.9	13.4	13.6	14.0	14.3
Percent of second stages	62.4	62.0	76.9	81.0	79.2
Percent of pie wasted	7.7	10.8	9.4	6.4	4.6
... conditional on second stage	12.3	17.4	12.2	7.8	5.8

Table 3: Summary of Bargaining Outcomes

5.1 Announcements in the First Stage

In the unrestricted design sessions, there is a wide range of demands made in the first stage. This heterogeneity is illustrated in figure 1. However, these announcements are concentrated on a small subset of demands. Table 4 provides summary statistics for the key announcements. Adding the induced 20 type in both *U1* and *U2* results in a significant increase in the proportion of announcements of 20 relative to the control sessions. Announcements of 10, which are complementary to the induced 20 demand, also show a significant increase in *U1*, and occur more frequently in that treatment than announcements of the induced 12 demand do in *U2*.

In the restricted design sessions, demands of 20 are the most frequent, followed by demands of 15. However, the proportion of demands of 15 is of a similar magnitude as 20 in *R3*. In both of these treatments there are few announcements of 18 (less than 10% of the time), and even fewer announcements (less than 5%) of 8 in *R4*.

Not only are demands of 18 infrequent in both *R3* and *R4*, but on average they lead to lower payoffs relative to demands of 15 and 20.³⁵ This result is shown in Table 5, which presents the average points payoff of subjects

³⁵ Significant at the 1% level using a one-sided t-test with robust standard errors, clustered at the session level.

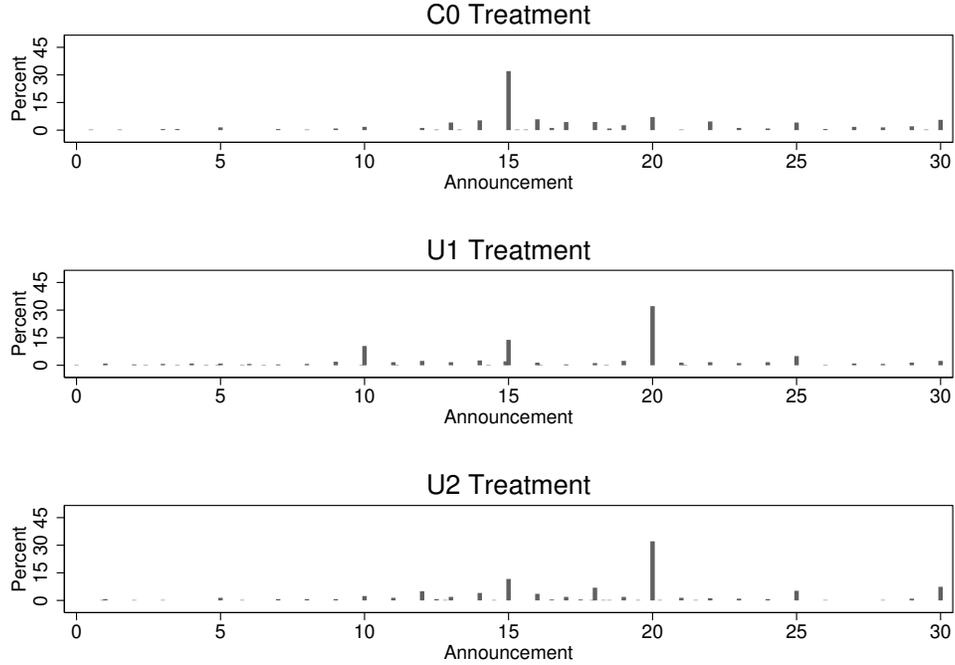


Figure 1: Subject Announcements in the Unrestricted Design

Treatment	Proportion of Announcements of											
	8		10		12		15		18		20	
<i>C0</i>	0.3	(1)	1.8	(6)	1.2	(4)	32.1	(109)	4.4	(15)	7.1	(24)
<i>U1</i>	0.7	(3)	10.5	(44)	2.4	(10)	13.8	(58)	1.2	(5)	32.1	(135)
p-value	0.386		0.008		0.365		0.201		0.150		0.035	
<i>U2</i>	0.7	(3)	2.4	(10)	5.0	(21)	11.7	(49)	6.9	(29)	32.1	(135)
p-value	0.386		0.606		0.152		0.169		0.264		0.010	
<i>R3</i>	.	(.)	.	(.)	.	(.)	44.9	(175)	7.9	(31)	47.2	(184)
<i>R4</i>	3.6	(13)	.	(.)	.	(.)	35.6	(128)	9.4	(34)	51.4	(185)

Table 4: Summary of Key Announcements

Number of observations are given in parentheses.
 P-values give the probability of a type I error in a two-sided t-test of the difference with the control (using robust standard errors clustered at the session level).

conditional on the announcement made.³⁶ In the sessions that were more likely to reach a second stage (treatments $U2$, $R3$ and $R4$), announcements of 20 lead to a greater payoff on average than announcements of 15. Making concessionary announcements leads to lower payoffs on average.³⁷ There are few demands that on average lead to more than half the pie.

Treatment	Average Payoff to Announcements of					
	8	10	12	15	18	20
$C0$	9.8	10.2	13.8	14.7	15.3	13.6
	.	(1.58)	(0.17)	(0.13)	(0.49)	(0.76)
$U1$	8.3	11.2	11.8	13.0	12.0	13.2
	(0.37)	(0.50)	(1.32)	(0.89)	(0.43)	(0.76)
$U2$	10.6	12.6	12.8	12.9	13.7	15.1
	(1.60)	(0.80)	(0.36)	(0.44)	(0.37)	(0.40)
$R3$.	.	.	12.6	11.6	14.3
				(0.32)	(0.34)	(0.33)
$R4$	9.8	.	.	13.7	12.8	14.6
	(0.49)			(0.43)	(0.57)	(0.31)

Table 5: Average Payoffs in Points

Parentheses give robust standard deviations, clustered at the session level.

5.2 Delay in the Second Stage

The expected delay in a subgame following the announcement of incompatible demands should be less than $\frac{1}{\lambda_{HL} + \lambda_{LH}}$ (where $\alpha_H \geq \alpha_L$), conditional on there eventually being agreement. Unlike equilibrium expected delay in a subgame, this bound gives a prediction for delay in a subgame that holds irrespective of the distribution of behavioural types. As a result, the analysis will not depend on either assumptions over or the estimation of this distribution.³⁸ Table 6 presents summary information on delay in the second

³⁶Note that these expected payoffs are only predicted to be equal if the announcements are part of the equilibrium mixing strategy *and* if all subjects are assumed to correspond to rational types. That this prediction depends on assuming all subjects are rational differs from the previous analysis of this subsection and the analysis that will follow in subsections 5.2 and 5.3, where the tested predictions are *not* dependent on such an assumption.

³⁷Both these results are significant at the 1% level using a one-sided t-test with robust standard errors, clustered at the session level.

³⁸As will become clear in the analysis that follows, this upper bound will be consistently violated by the data. As a consequence, the same conclusions would hold for equilibrium expected delay, irrespective of the assumptions or estimates used to calculate them.

stage, in particular comparing the mean delay to this upper bound for subgames with at least 10 observations.³⁹ The observed delays are significantly longer than their respective upper bounds for all treatments (significant at the 1% level).⁴⁰ This result is emphasized by looking at the last column of table 6, which gives the average ratio of mean delay to upper bound. The smallest value of this ratio is 3.78, which is observed in $R4$. This means that on average the mean delay in any given subgame of $R4$ is more than triple its respective bound.

Treatment	Subgame [§]		Obs		Delay (Seconds)			
	α_L	α_H	Freq	%	Mean	Bound	Ratio ^{§§}	
$C0$	<i>All</i>		87		202.6			5.00
$U1$	<i>All</i>		105		263.9			8.44
	15	20	18	17.1	80.8	20.0		
	20	20	22	21.0	411.4	50.0		
$U2$	<i>All</i>		115		118.6			8.11
	15	20	12	10.4	59.2	20.0		
	20	20	14	12.2	70.9	50.0		
$R3$	<i>All</i>		124		87.0			4.86
	15	20	78	62.9	63.6	20.0		
	18	20	14	11.3	151.1	36.4		
	20	20	26	21.0	128.7	50.0		
$R4$	<i>All</i>		103		64.0			3.78
	15	20	46	44.7	22.2	20.0		
	18	20	11	10.7	114.0	36.4		
	20	20	36	35.0	111.3	50.0		

Table 6: Second-Stage Delay

[§] Only subgames with at least 10 observations reported.

^{§§} Weighted average of mean delay divided by bound.

Contrasting the performance across designs, mean delay comes closer to the bounds dictated by the model in the restricted design: the delay to bound ratio is significantly smaller at 4.37, compared to 7.34 in the unrestricted design (significant at the 5% level). However, restricting attention in the unrestricted sessions to subgames involving announcement of 15 and

³⁹Unless otherwise stated, only matches where both players are human subjects are included in the second stage analysis.

⁴⁰Results from a Wald test on the coefficients of a regression of the ratio of delay to upper bound on treatment dummies. Robust standard errors, clustered at the session level, were used. Throughout this subsection, statistical significance is established using analogous tests. See section 3.3 of the online appendix for further details.

20, results in a more favorable comparison with the restricted design. Indeed, the difference between this sub-sample of the unrestricted design and the restricted design is no longer significant. Despite the indications of a better performance in these explicit subgames, and the restricted design in general, the observed delay is still consistently above that predicted by the model, suggesting that subjects are engaging in excessively costly second-stage delays.⁴¹

5.3 Concession in the Second Stage

The model imposes structure on the pattern of concession behavior. Rational subjects will either concede instantly, if the demand made by the other player is not mimicked in the first stage,⁴² or employ a mixed strategy over the time to concession.⁴³ In the latter case, excluding the possibility of initial concession, the player who made the lower demand is more likely to concede, but should never concede initially.

The top panel of figure 2 presents the probability that either player concedes for asymmetric subgames with at least 10 observations in the unrestricted design.⁴⁴ Overall, it is not clear that interior concession behavior is in line with the predictions of the model, in particular for the control sessions where the subject making the higher announcement is observed to concede more often. It is not possible to reject the null hypothesis that the probability of concession by the lower announcer is less than or equal to the probability of concession by the higher announcer.⁴⁵ However, in both *U1* and *U2*, concession in the 15 – 20 subgames is consistent with the predicted pattern: using this sub-sample, the probability of the lower announcer con-

⁴¹See section 3.3 of the online appendix for further evidence of excessive delays.

⁴²In equilibrium the probability the other player would be irrational is 1, meaning there is no possible benefit to holding out for any length of time.

⁴³The mixed strategy has the form of arriving with a constant hazard rate with the player making the smaller demand conceding with the greater rate. These probabilities of conceding are unconditional on knowing whether the player is rational or not.

⁴⁴Attention is restricted to asymmetric subgames since only these have a theoretical prediction.

⁴⁵Results from a Wald test on the coefficients of a regression of an indicator variable for interior concession by the lower announcer on design dummies. Robust standard errors, clustered at the session level, were used. Throughout this subsection, statistical significance is established using analogous tests. See section 3.4 of the online appendix for further details.

ceding is strictly larger than that of the higher announcer (significant at the 1% level). With respect to initial concession behavior, there is little evidence that players concede initially in any of the unrestricted design treatments,⁴⁶ which is surprising since a non-aggressive type was induced in *U2* and a complementary type emerged in *U1*.

The bottom panel of figure 2 presents the data on concession behavior for the restricted design. First, the model predicts that the player who made the higher announcement should concede initially in *R4*, but should never do so in *R3*. Second, excluding initial concession, the player making the lower announcement should be more likely to concede in both treatments. These predictions are observed in the figure, particularly for 15 – 20 subgames. In *R3*, the lower announcer is observed conceding more often in both the initial and interior panels. In *R4*, while the lower announcer continues to concede more often in the interior panel, it is the higher announcer that is observed to concede initially more often, as predicted. However, while these observations are all in the direction predicted by the model, only the initial concession comparative static is statistically significant (at the 1% level).

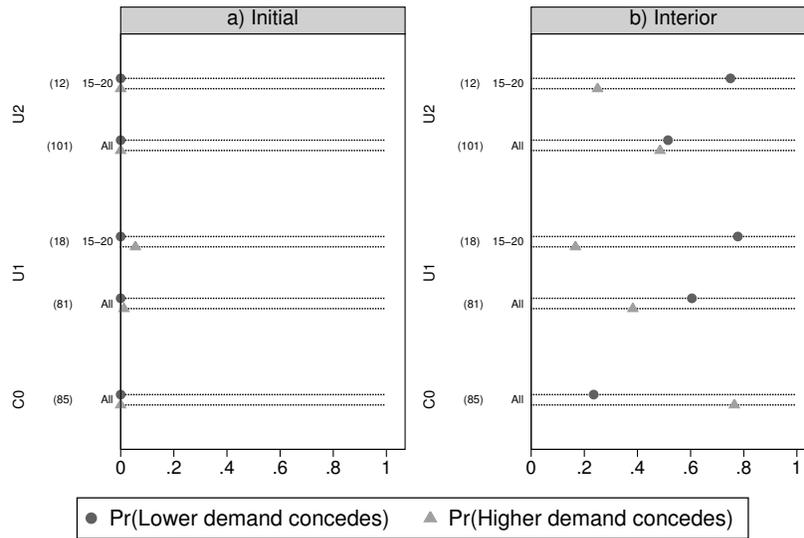
6 Discussion

Bargaining Outcomes

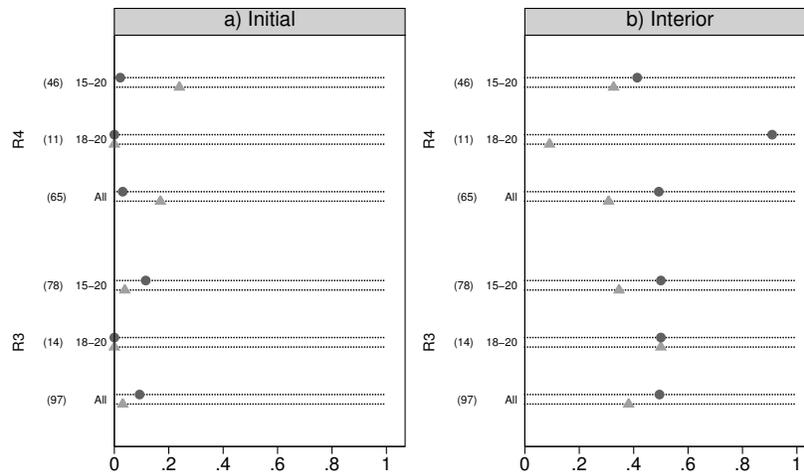
As a benchmark, the results from the control sessions share both similarities and differences with results from prior bargaining experiments. In our experiment, about one-third of demands are for half of the pie. Although this is not a negligible number, it is much less than the 100% observed in the comparable treatment of *Roth and Malouf (1979)*. We do observe more conflict than is typical, with 62.4% of first-stage demands being incompatible and resulting in some delay. In a review of prior experiments discussed in *Roth (1995)*, inefficient outcomes in comparable ultimatum and shrinking-pie games are observed at most 42% of the time (based on numbers from *Forsythe et al. 1991*). However, since incompatible first-stage demands do not imply outright disagreement in our environment, another interesting comparison is

⁴⁶In the analysis presented here, initial concession is represented by concession within the first 2 seconds. Interior concession is represented by concession outside the first 2 seconds. Results are robust to small changes in the specific cutoff; see section 3.4 of the online appendix for further details.

Unrestricted Design



Restricted Design



Asymmetric subgames with at least 10 observations. Parentheses give number of observations.
 Initial concession is represented by concession within the first 2 seconds.
 Interior concession is represented by concession after the first 2 seconds.

Figure 2: Concession Behavior in Asymmetric Subgames

the fraction of the pie that is destroyed. We find 7.7% of the pie is lost to delay, a percentage that is comparable to these prior studies. To summarize, subjects in our implementation regularly adopted adversarial positions during the bargaining process, but on average avoided wasting too much of the available pie.

Recognizing the Role of Reputation

The first prediction the experimental design sought to investigate was whether subjects recognize the role of reputation in this environment. To this end, the unrestricted design aimed to test if players would mimic obstinate types while imposing the minimum possible restriction on subject announcement behavior. The comparison between first-stage announcements in the control and those in sessions with induced types provides support for this prediction: following the introduction of an experimentally induced 20-type, there is a substantial increase in demands of 20. In moving away from the dominant announcement of 15 from the control, subjects display a willingness to give up an allocation that is not only equitable and efficient, but also the strong 50-50 norm of this symmetric bargaining environment.

This observation is reminiscent of results from earlier unstructured bargaining experiments. For example, in experiments with bargaining over lottery points, *Roth et al.* (1981) found that when one player has a lower prize in the final lottery than the other - and this is common knowledge to both parties - subjects would tend to negotiate over an equal expected-value allocation or an equal lottery-points allocation. The former allocation requires giving more lottery points to the disadvantaged player to equate expected payoffs, while the latter implies a higher expected value for the advantaged player. Significantly, the player with the higher lottery payoff was more likely to suggest the equal lottery-points allocation, thus using an available notion of fairness in a strategic manner. The differences with our implementation are that there is a precise theoretical prediction concerning the strategies bargainers should take, and the behavior subjects are predicted to imitate cannot be interpreted as fair.

Our findings also parallel a growing body of research that investigates the extent to which subjects will behave in a more self-interested manner when they are given a means by which to hide their intentions. For example, *Andreoni and Bernheim* (2009) find that the observability of the dictator's choice affects the frequency of equal splits in the dictator game. They show

that, when there is the possibility that the dictator’s choice might be switched for another, unequal allocation, the dictator is more likely to offer this unequal split than when such an intervention is not possible. *Dana et al.* (2006, 2007) document similar behavior with their variants of the dictator game. In our experiments, we find comparable behavior in a strategic environment, with subjects using the presence of computer players to move away from the prevalent 50-50 norm, even though such an action may lead to conflict in a second stage. The fact that the environment is a strategic, as opposed to an individual decision, is a substantive difference both in terms of complexity and in terms of identifying the important behavioral tendencies. Further, there is no possible conflict that could result from subjects taking the more self-interested action in these dictator games. This is not the case with the stylized bargaining game where an aggressive demand will most likely lead to conflict, making it far from clear that subjects would have been willing to follow such strategies.

One aspect of our results that differs substantially from the early studies of reputation, in particular *Camerer and Weigelt* (1988) and *Jung et al.* (1994), is that - to the extent that they found results in line with the sequential equilibrium prediction - many repetitions or experienced subjects were needed. In our experiment, subjects displayed mimicking behavior from the early periods. In their study of one-sided reputation, *Grosskopf and Sarin* (forthcoming) also find behavior consistent with reputation building without needing many repetitions.⁴⁷ However, their results suggest that the theoretical predictions of reputation effects are more likely to be borne out when such effects are aligned with the prevalent other-regarding preferences of the environment; for example, when the reputation effect requires an honest action rather than a dishonest one. This is in contrast to our results, where subjects are observed building reputation as a type demanding strictly more than the 50-50 split (the prevalent norm), despite this split always being an available alternative.

Further Analysis of Announcement Strategies

Several other features of subject behavior warrant further discussion. First, in the *U1* sessions, not only is there an increase in 20 announcements, but

⁴⁷A common feature of these two designs is the use of computer players, rather than subjects with dominant strategies, to induce reputational types.

also the introduction of two 20 type computers leads to an increase in demands of 10.⁴⁸ Such demands are complementary to the induced type in that they ensure that a second stage will not be reached. Note that a rational player of the model would strictly prefer to announce 15 rather than 10, even if they knew with probability one that a player announcing 20 is a behavioral type: should they encounter an announcement of 20 they could always concede instantly, thus giving themselves a payoff of 10. This alternative strategy would not involve any delay and guarantee at least as high a payoff. Consequently, the presence of complementary types suggests that some subjects are willing to accept an inferior payoff in order to avoid the process of negotiation altogether. Although the model used in this paper does not predict an increase in such a type, the coexistence of complementary pairs, such as 10 and 20, is an outcome of the evolutionary stability analysis of *Abreu and Sethi* (2003).⁴⁹

Second, there is support for the conjecture that subjects mimic the most aggressive demand too often. As shown in table 4, there are significantly more announcements of 20 than 18 in the *R3* treatment. This observation can only be rationalized by the model if there are more 20 types in the population than 18 types. Ex-ante, since 15 is the fair split, it seems reasonable to have expected more 15 types in the population than either 18 or 20 types. However, there does not appear to be any reason to expect more 20 types than 18 types.

Third, the inclusion of a concessionary announcement of 8 in the *R4* treatment provides a marker for the size of any potential demand-induced effects. Announcements of 8 should never be mimicked by rational subjects. As table 4 shows, such demands were made less than 5% of the time, suggesting demand-induced effects are minimal. A related result is that demands of 12 in the *U2* treatment are of the same order, which is in line with the basic comparative static as discussed in section 4, reinforcing the evidence that demand-induced effects are of a second order.

⁴⁸Such complementary types did not emerge in significant numbers the *U2* treatment.

⁴⁹While it is unlikely that evolution of behavior is being observed within the short time frame of these experiments, it is interesting to note that such a model captures something that is observed in the laboratory.

Role of an Explicit Set of Types

In both the *U1* and *U2* treatments, rational subjects knew for certain that demands of 20 could be from an irrational player. It is also plausible that they would expect some subjects to insist on getting half of the pie, a demand of 15, since it is such a salient norm. In these treatments, it is when these two demands meet that second-stage concession behavior comes closer to the predicted pattern. Under the restricted design, where the set of behavioral types is common knowledge, second-stage behavior is closer to the predictions of the model (see table 6 and figure 2). These observations combined suggest that uncertainty over the set of behavioral types could be responsible for some of the observed departures from the theory in the unrestricted design.

More generally, it is worth stressing that the quantitative details of the sequential equilibrium predictions are quite complicated. Subjects are required to employ not only a mixed strategy over first-stage announcements, but also a mixed strategy over concession time, all calibrated to ensure their (rational-type) opponent is indifferent over their respective choices. A number of experimental studies (for example *Shachat* 2002) have noted that subjects find mixed strategies difficult to implement. *Camerer and Weigelt* (1988) and *Jung et al.* (1994), who find support for the sequential equilibrium in their studies of reputation, also report evidence that subjects do not mix in the correct proportions.⁵⁰ Hence, it is not surprising to find deviations with respect to the finer details of the equilibrium predictions, but systematic deviations, such as those identified, are nonetheless informative.

While part of the identified deviations can be accounted for by uncertainty over the set of behavioral types, more is needed for the model to accommodate the rest, such as making more demanding announcements too often. One extreme assumption of the model is that behavioral types *never* concede; in line with the model, this is how our computer players are programmed to behave. However, it seems unreasonable to suppose that subjects that are strategically unresponsive would be so inflexible as to never concede no matter how much the pie has already shrunk. Including the possibility that obstinate types from the subject population might eventually concede results in both greater delay, conditional on eventual agreement, and an increase in more

⁵⁰Delaying concession for too long has also been observed in all pay auctions, a setting with arguably a simpler strategy for subjects to implement: rather than mixing over concession times, they play a pure strategy following the random draw from their cost function.

demanding announcements. Thus, a minor modification to the theory would bring the finer details of the equilibrium closer to the observed behavior.⁵¹

7 Conclusions

In implementing the stylized model of *Abreu and Gul* (2000), the experiments presented in this paper investigate behavior in a setting that underlies a general class of bargaining environments. In using the stylized game, the design has the advantage of eliminating a potential confound on bargaining power, namely proposer power, that arises in other protocols. The experiments focus on the role played by reputation, while providing the structure needed for testable predictions on players' strategies. At the core of these predictions is the need to mimic the demands of obstinate players: for a player to do otherwise would reveal rationality and lead to a weak bargaining position. We find clear evidence that subjects recognize this role for reputation and mimic the induced obstinate types. We also find evidence for the presence of complementary types, something not predicted by this model, but consistent with an evolutionary view of the environment. Complementary types will acquiesce to credibly obstinate demands, rather than go through the negotiation process to reach an agreement.

While the qualitative predictions of the first stage of the model are borne out in the laboratory, some of the finer details of the sequential equilibrium predictions are not: most notably, the tendency to make more demanding announcements too often and to remain in the concession stage too long. These deviations are not surprising given the complexities of the sequential equilibrium and prior experimental evidence demonstrating the difficulties subjects have in implementing mixed strategies. However, what is surprising is that these deviations are all in the direction of more conflict, inequality and inefficiency, which is in contrast to much of the bargaining literature that generally reports deviations that work in the opposite direction. Furthermore, a minor modification to the assumed behavior of unresponsive types would result in predictions with more demanding first-stage announcements and greater delay. Finally, the performance of the model improved under the restricted design, with second-stage behavior coming closer to the

⁵¹This modification is in the spirit of the perturbation of the standard alternating-offers bargaining given in *Binmore and Swierzbinski* (2006). See section 4 of the online appendix for further details.

predicted pattern. These observations indicate the importance of having an explicit set of behavioral types for empirical applications, as well as the theory. Results from the second stage of the unrestricted design suggest that, should the set of behavioral types not be evident to bargaining parties, the process by which such behavioral play becomes credible needs to be modeled directly.

To summarize, the evidence presented in this paper suggests that individuals behave strategically in bargaining environments, even when such behavior is in opposition to the strong norms of equality and efficiency. At first glance, this observation appears to be in sharp contrast to the results from other bargaining experiments, where subjects are generally only observed to behave strategically if their behavior coincides with a plausible norm of fairness. However, our results are consistent with an increasing body of research that has investigated the role of intentions in decision making. In mimicking the induced types, we observe subjects willing to take an adversarial stance in their first-stage announcements, despite this being contrary to the prevalent other-regarding preference and possibly leading to conflict in a subsequent round. Undoubtedly, more research is needed to understand further the extent to which the *Abreu and Gul* (2000) model can be used to predict behavior in bargaining situations. Nonetheless, our results indicate that this two-sided reputation approach is an appropriate tool for investigating bargaining behavior, particularly in environments where the possibilities for norm-driven or obstinate play are transparent to all actors.

References

- Abreu, D., and F. Gul, Bargaining and reputation, *Econometrica*, 68, 85–117, 2000.
- Abreu, D., and R. Sethi, Evolutionary stability in a reputational model of bargaining, *Games and Economic Behavior*, 44, 195–216, 2003.
- Andreoni, J., and B. D. Bernheim, Social image and the 50-50 norm: A theoretical and experimental analysis of audience effects, *Econometrica*, 77(5), 1607–1636, 2009.
- Andreoni, J., and J. H. Miller, Rational cooperation in the finitely repeated

- prisoner's dilemma: Experimental evidence, *Economic Journal*, pp. 570–585, 1993.
- Atakan, A. E., and M. Ekmekci, Bargaining and reputation in search markets, *Discussion paper*, 2010.
- Babcock, L., and G. Loewenstein, Explaining bargaining impasse: The role of self-serving biases, *The Journal of Economic Perspectives*, 11(1), 109–126, 1997.
- Baron, D. P., Private politics, *Journal of Economics & Management Strategy*, 12(1), 31–66, 2003.
- Binmore, K., and J. Swierzbinski, A little behavioralism can go a long way, *Does Game Theory Work? The Bargaining Challenge*, 8, 257–276, 2006.
- Bolton, G. E., and A. Ockenfels, Erc: A theory of equity, reciprocity, and competition, *The American Economic Review*, 90, 2000.
- Brandts, J., and N. Figueras, An exploration of reputation formation in experimental games, *Journal of Economic Behavior & Organization*, 50, 89–115, 2003.
- Bulow, J., and P. Klemperer, The generalized war of attrition, *The American Economic Review*, 89, 175–189, 1999.
- Camerer, C., and K. Weigelt, Experimental tests of a sequential equilibrium reputation model, *Econometrica*, 56, 1–36, 1988.
- Charness, G., and M. Rabin, Understanding social preferences with simple tests, *The Quarterly Journal of Economics*, 117, 817–869, 2002.
- Chatterjee, K., and L. Samuelson, Bargaining with two-sided incomplete information: An infinite horizon model with alternating offers, *The Review of Economic Studies*, 54, 175–192, 1987.
- Chatterjee, K., and L. Samuelson, Bargaining under two-sided incomplete information: The unrestricted offers case, *Operations Research*, 36, 605–618, 1988.

- Cooper, D. C., and J. H. Kagel, Other regarding preferences: A survey of experimental results, *Handbook of Experimental Economics*, 2, forthcoming.
- Crawford, V. P., A theory of disagreement in bargaining, *Econometrica*, 50, 607–637, 1982.
- Damiano, E., H. Li, and W. Suen, Optimal deadlines for agreements, *Theoretical Economics*, 7(2), 357–393, 2012.
- Dana, J., D. M. Cain, and R. M. Dawes, What you don't know won't hurt me: Costly (but quiet) exit in dictator games, *Organizational Behavior and Human Decision Processes*, 100, 193–201, 2006.
- Dana, J., R. Weber, and J. Kuang, Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness, *Economic Theory*, 33, 67–80, 2007.
- Falk, A., E. Fehr, and U. Fischbacher, On the nature of fair behavior, *Economic Inquiry*, 41, 20–26, 2003.
- Fehr, E., and K. M. Schmidt, A theory of fairness, competition, and cooperation, *Quarterly Journal of Economics*, 114, 817–868, 1999.
- Forsythe, R., J. Kennan, and B. Sopher, An experimental analysis of strikes in bargaining games with One-Sided private information, *The American Economic Review*, 81(1), 253–278, 1991.
- Grosskopf, B., and R. Sarin, Is reputation good or bad? an experiment, *The American Economic Review*, forthcoming.
- Hörisch, H., and O. Kirchkamp, Less fighting than expected - experiments with wars of attrition and all-pay auctions, *Public Choice*, 144(1), 347–367, 2010.
- Jung, Y. J., J. H. Kagel, and D. Levin, On the existence of predatory pricing: An experimental study of reputation and entry deterrence in the chain-store game, *The RAND Journal of Economics*, 25, 72–93, 1994.
- Kambe, S., Bargaining with imperfect commitment, *Games and Economic Behavior*, 28, 217–237, 1999.

- Kwiek, M., Reputation and cooperation in the repeated second-price auctions, *Journal of the European Economic Association*, 9(5), 982–1001, 2011.
- McKelvey, R., and T. Palfrey, *The holdout game: an experimental study of an infinitely repeated game with two-sided incomplete information*, in “*Social Choice, Welfare, and Ethics*” (W. Barnett, H. Moulin, M. Salles, and N. Schofield, Eds.), Cambridge University Press, Cambridge, 1995.
- Muthoo, A., *Bargaining Theory with Applications*, Cambridge University Press, 1999.
- Myerson, R. B., *Game theory: analysis of conflict*, Harvard University Press, 1991.
- Neral, J., and J. Ochs, The sequential equilibrium theory of reputation building: A further test, *Econometrica*, 60, 1151–1169, 1992.
- Ochs, J., and A. E. Roth, An experimental study of sequential bargaining, *The American Economic Review*, 79, 1989.
- Özyurt, S., Searching a bargain: power of strategic commitment, 2011.
- Povel, P., Multiple banking as a commitment not to rescue, *Research in Finance*(21), 175–199, 2005.
- Roth, A. E., Bargaining experiments, *Handbook of Experimental Economics*, pp. 253–348, 1995.
- Roth, A. E., and M. W. K. Malouf, Game-theoretic models and the role of information in bargaining, *Psychological Review*, 86, 574–594, 1979.
- Roth, A. E., and J. K. Murnighan, The role of information in bargaining: An experimental study, *Econometrica*, 50, 1123–1142, 1982.
- Roth, A. E., and F. Schoumaker, Expectations and reputations in bargaining: An experimental study, *The American Economic Review*, 73, 362–372, 1983.
- Roth, A. E., M. W. K. Malouf, and J. K. Murnighan, Sociological versus strategic factors in bargaining, *Journal of Economic Behavior & Organization*, 2, 153–177, 1981.

Rubinstein, A., Perfect equilibrium in a bargaining model, *Econometrica*, 50, 97–109, 1982.

Schelling, T. C., The strategy of conflict, *Cambridge, Mass*, 1960.

Shachat, J. M., Mixed strategy play and the minimax hypothesis, *Journal of Economic Theory*, 104, 189–226, 2002.

Wolitzky, A., Reputational bargaining with minimal knowledge of rationality, *Econometrica*, forthcoming.