Review of Economic Studies (2015) **82**, 608–631 doi:10.1093/restud/rdu029

Advance access publication 19 September 2014

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

MATTHEW EMBREY
Maastricht University

GUILLAUME R. FRÉCHETTE
New York University

and

STEVEN F. LEHRER
Queen's University and NBER

First version received September 2012*; final version accepted July* 2014 (*Eds.*)

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 acquiesce to induced behavioural 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.

Key words: Bargaining, reputation, Experiment

JEL Codes: C78, C91, D82

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

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

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 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 article, 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 behaviour: first, there is more conflict than in previous bargaining studies; second, subjects fight for their demands longer than predicted, and third, while there is clear evidence that subjects recognize the role of reputation in this environment, they tend to be more demanding in their announcements than predicted. In all cases, these observations result in less equality and less efficiency. This contrasts with much of the prior experimental literature on bargaining, where increasing equality and efficiency are the two primary 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.

2. In 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 article 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.

3. The 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.

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

5. The 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.

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 introduced in Abreu and Gul (2000) would also make disadvantageous counter-offers in such environments. An alternative 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 (2012) and Atakan and Ekmekci (2014) — 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 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. 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 article presents analyses from six treatments that differ in two dimensions. The first is the demands that can be made: either any feasible demand (the unrestricted design) or only a predetermined subset (the restricted design). The second is the introduction of computer players that follow a pre-specified strategy (referred to as induced behavioral types).[8] The unrestricted design aims to determine if the subjects recognize the role of posturing in bargaining.[9] This will be the first, and in our view the key, comparative static: do the subjects mimic the induced

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

7. See 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).

8. In each design, there is one treatment with no induced types and two treatments with induced types; the treatments with induced types differ in the number and behavior of these types. We 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), Anderhub, Engelmann, and Güth (2002) and Grosskopf and Sarin (2010), among others.

9. In this model, such comparative static predictions depend on the specific distribution of behavioral types. One of the innovations of this article is to recover a distribution of behavioral types consistent with the behavior in the treatment without induced behavioral types (*i.e.* among the subjects) to confirm that the comparative static of interest applies to

behavioral types when rational agents are predicted to do so. 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, such as the 50–50 norm. Under the restricted design, subject demands are restricted to a set with just three or four alternatives. In all but one baseline treatment, this set corresponds to the demands 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 behavioral types: across treatments there is an increase in subject demands that correspond to induced types predicted to be mimicked. In addition, the experiment reveals 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.

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 sessions with induced types is mostly consistent with this hypothesis, although some disparities do persist: subjects appear to be disproportionately mimicking the most demanding type in the first stage, and delays in the second stage remain longer than predicted. We propose a minor modification to the assumed behavior of obstinate types that results in predictions with more demanding first-stage announcements and greater delay, thus accommodating the identified disparities.

The article 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 article 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 in bargaining regarding the establishment of focal points, what

our data. This novel combination of structural estimation and experimental design generates testable predictions in a model that is otherwise too flexible to generate such predictions without stronger ex-ante assumptions on the behavior of subjects. In addition, the estimation does not rely on the addition of an arbitrary noise structure: any demand observed can be generated with some strictly positive probability within the original model for appropriately selected parameters.

10. To 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).

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] Using a repeated trust game, Wilson and Eckel (2006) present evidence that high levels of trust may simply be an investment in a subject's own reputation. In a related market environment, Bohnet *et al.* (2005) demonstrate that providing feedback, which allows players to observe their partner's history, improves the functioning of the market since participants are found to learn the mechanics of reputation building by imitation. Anderhub, Engelmann, and Güth (2002) use computers to induce the probability of facing a second mover who always reciprocate in a trust game. They conclude that reputation formation provides a reasonable account of their data.

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 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 analogue 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

11. Studies 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.

12. This 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.

13. For 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.

14. See Neral and Ochs (1992), Brandts and Figueras (2003), and Grosskopf and Sarin (2010). Jung *et al.* (1994) use a similar design but test the chain store game. Tests of the repeated trust game also include Engle-Warnick and Slonim (2004).

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

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 the 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 demand α^i (*i.e.* the faction of the pie they would like). If the two demands are compatible (*i.e.* $\alpha^1+\alpha^2 \leq 1$) then the game ends immediately.[17] If the two demands 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 can choose to accept (*i.e.* concede) or hold out. If player i concedes, then player i receives $1-\alpha^j$, 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 that player's 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 $C := \{\alpha_1,...,\alpha_K\}$ as the set of behavioral types, with $\alpha_i < \alpha_{i+1}$, for $i=1,...,K-1$, and $\alpha_K \geq \frac{1}{2}$. 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 $k=1,...,K$. The probability that a player is rational is denoted by $z_0 = 1-\sum_{k=1}^{K} z_k$. This distribution over types is summarized by the vector $z=\{z_1,...,z_K\}$. Finally, a behavioral type is defined as *aggressive* if it is incompatible with all other behavioral types in C and weakly incompatible with itself.

3.1. *Equilibrium behaviour*

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

$$\overline{\pi}_k = \frac{z_k}{z_k + z_0\mu_k}. \tag{1}$$

16. What is contained here is implicit in Abreu and Gul (2000), whose stylized game differs slightly to our experimental implementation in that it is not symmetric — in particular, initial announcements are made sequentially and players can have different sets of behavioral types (with different distributions over these sets). Some results for equilibrium announcements are drawn from Abreu and Sethi (2003), where the authors use the symmetric version of the model, as implemented here — that is, players make initial announcements simultaneously and have a common set of behavioral types (and distribution over this set).

17. If 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.

In equilibrium, with a general set C, rational players will employ a mixed strategy over announcing different types in the 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 $\{\alpha_R,...,\alpha_K\}$, where $1 \leq R < K$. Ensuring that rational players are indifferent between announcing any α_k (for $k=R,...,K$), 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]

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 by[20]

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

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

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

and $c_{kl} = \overline{\pi}_k e^{\lambda_{kl}T_0}$ 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.* without knowing if the α_k-player is 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-player will 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.

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 (C)

18. Both the support of μ and the expected payoff are uniquely determined by the parameters (C,z). However, calculating these requires solving this system of non-linear equations, which only has a numerical solution except for trivial cases. Section 1.7 of the Supplementary Appendix contains details of the numerical strategy used to solve this system of equations for a general set of behavioral types. The Supplementary Appendix is available at both the Review of Economic Studies online and the individual author's websites.

19. Note 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.

20. So 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 without knowing whether the opponent is rational or not. Equation (2) ensures this indifference holds.

and the distribution over these types (z), there are a number of features of equilibrium behavior that hold more generally.[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.
3. Aggressive types are always mimicked in equilibrium.
4. If the set C contains *only* aggressive types, then

$$\alpha_k > \alpha_l \Longrightarrow \mu_k < \left(\frac{z_k}{z_l}\right) \cdot \mu_l.$$

That is, if the less demanding type is at least as probable as the more demanding type, then rational players will mimic the more demanding announcement less often.

5. If the set C contains a type that is compatible with all other types in C (*i.e.*, $\alpha + \alpha_i \leq 1$, for all $\alpha_i \in C$), then this type is never mimicked by rational players.

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]

6. If $\alpha_L = \alpha_H$ or if both α_L and α_H are aggressive, there is no initial concession by either player.
7. 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.
8. Since $\alpha_L < \alpha_H$ implies that $\lambda_{LH} < \lambda_{HL}$, the more demanding α_H-announcers concede at a slower rate than the α_L-announcers for interior concession (without conditioning on the players being rational or not). Consequently, it is more likely that an α_L-announcer makes an interior concession than an α_H-announcer.
9. The upper bound on the average (mean) delay, given that agreement is eventually reached, is $\frac{1}{\lambda_{LH} + \lambda_{HL}}$.

4. EXPERIMENTAL DESIGN

In order to implement the stylized game in the laboratory, 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

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

22. Throughout the article, 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.

23. An 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

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 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.[26]

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, in *U1* there would be only announcements of 20, and in *U2* there would be only announcements of 12 and 20.[27]

However, given the evidence from bargaining experiments for norm-driven behavior, it is reasonable to postulate that not all subjects who enter the laboratory correspond to a rational

actually do in this situation, computer players were used instead to maintain control over the environment. Furthermore, the alternative would generate concerns regarding what other subjects believe about how these subjects would choose to behave.

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

25. See sections 2.1 and 2.2 of the Supplementary Appendix for sample instructions and screen shots.

26. Subjects were recruited for up to 2 hours. In all sessions, all 15 periods of bargaining were completed within the scheduled time.

27. See Tables 1 and 2 of the Supplementary Appendix for precise numerical predictions. When 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.

TABLE 1
Summary of all sessions

Treatment		Session 1	2	3	4	5	Total
C0	Subjects	12	12	10	10	14	58
	Computers	.	.	.	.	.	
	Earnings ($)	25.0	23.3	23.0	19.0	20.7	22.2
U1	Subjects	14	14	14	14	14	70
	Computers	20, 20	20, 20	20, 20	20, 20	20, 20	
	Earnings ($)	20.7	23.6	22.1	20.7	23.6	22.1
U2	Subjects	14	14	14	14	14	70
	Computers	12, 20	12, 20	12, 20	12, 20	12, 20	
	Earnings ($)	22.1	25.0	27.9	23.6	29.3	25.6
R0	Subjects	10	16	16	16	12	70
	Computers	.	.	.	.	.	
	Earnings ($)	25.0	22.5	18.8	22.5	21.7	21.9
R3	Subjects	13	13	13	13	13	65
	Computers	15, 18, 20	15, 18, 20	15, 18, 20	15, 18, 20	15, 18, 20	
	Earnings ($)	24.2	24.2	19.6	24.2	24.2	23.3
R4	Subjects	12	12	12	12	12	60
	Computers	8, 15, 18, 20	8, 15, 18, 20	8, 15, 18, 20	8, 15, 18, 20	8, 15, 18, 20	
	Earnings ($)	25.0	25.0	20.0	23.3	25.0	23.7

type of the model.[28] Since the exact predictions of the model are dependent on the distribution of behavioral types, the distribution of potential behavioral types among the population of subjects must also be considered, along with any induced types. Despite this complication, for many assumptions over such types moving from the control to either *U1* or *U2* would imply an increase in the number of 20 demands, while the demand of the 12-computer would either not be mimicked by rational subjects at all, or rarely so.

Table 2 gives the predictions of the model for several example scenarios that include the possibility of subjects being irrational, which should be viewed as complementary illustrations of this basic comparative static. First, scenario A considers the possibility that on average one out of 13 subjects brought into the laboratory acts like a 15-type.[29] That is, they always demand the 50–50 split and would never accept less. As the last set of columns of the table shows, while the probability of observing an announcement of 12 by a subject is not predicted to change across the treatments, moving from the control to either *U1* or *U2* should result in an increase in demands of 20.[30]

Clearly it is unrealistic to assume 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,

28. This 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.

29. In either *U1* or *U2*, if on average one out of 13 human players (*i.e.* subjects) were a 15-type, then for a rational subject the probability of being matched to this 15-type subject would be $1/15$. This is because, along with the 13 other human players, there are also two computer players in the matching pool. In the control treatment, which does not include the two computer players, this probability would be $\frac{1}{13}$. For this reason, the various rationality scenarios consider different multiples of $\frac{1}{13}$ for the distribution of types in the subject population.

30. This 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.

TABLE 2
Example rationality scenarios

Treatment		z_0	z 12	15	20	μ 12	15	20	Prob. Observing§ 12	15	20
Scenario A[1]											
C0	Assumed	92.3	0.0	7.7	0.0	0.0	100.0	0.0	0.0	100.0	0.0
U1	Predicted	80.0	0.0	6.7	13.3	0.0	53.4	46.6	0.0	57.0	43.0
U2	Predicted	80.0	6.7	6.7	6.7	0.0	69.0	31.0	0.0	71.4	28.6
Scenario B[2]											
C0	Estimated	7.7	1.1	30.4	9.7	0.0	9.9	23.2	1.1	31.2	11.5
U1	Predicted	6.7	0.9	26.4	21.7	0.0	8.9	41.3	1.1	31.1	12.9
U2	Predicted	6.7	7.6	26.4	15.1	0.0	0.0	38.9	1.1	30.4	12.7
Scenario C[3]											
C0	Estimated	61.5	1.1	7.1	3.6	0.0	39.4	12.7	1.1	31.4	11.4
U1	Predicted	53.3	0.9	6.2	16.5	0.0	24.2	40.9	1.1	22.	28.8
U2	Predicted	53.3	7.6	6.2	9.8	0.0	29.3	31.3	1.1	25.2	22.9
Scenario D[4]											
C0	Estimated	92.3	0.5	1.2	1.0	0.0	69.2	15.4	0.5	65.1	15.2
U1	Predicted	80.0	0.4	1.0	14.2	5.5	13.4	75.4	5.5	13.6	70.6
U2	Predicted	80.0	7.1	1.0	7.5	17.2	18.5	57.1	16.3	18.3	53.7

§ Probability of observing an announcement, excluding those made by computer players.
All values in the table are probabilities represented as percentages (for reference, $\frac{1}{13} \approx 7.7\%$ and $\frac{1}{15} \approx 6.7\%$). The μ columns do not sum to one in scenarios B, C and D since $\{12, 15, 20\}$ does not include the entire support of μ.
[1] z_{15} assumed so that on average 1 out of 13 human players is a 15-type.
[2] z estimated using announcement data, assuming $z_0 = \frac{1}{13}$.
[3] z estimated using announcement data, assuming $z_0 = \frac{8}{13}$.
[4] z estimated using announcement data, assuming $z_0 = \frac{12}{13}$.

identified up to a given value for the probability of being matched to a rational player. Indeed the model generates a strictly positive probability of observing any demand for appropriately selected parameters, allowing us to recover the implied distribution of behavioral types. Using these estimates, the implied probability of demands of 12 and 20 in *U1* and *U2* can be predicted, and the basic comparative static confirmed empirically for our specific data. The results of this analysis are reported in Table 2 as scenarios B, C, and D respectively, where the first assumes a small probability of being matched to a rational subject, the third a large probability and the second something in between. The basic comparative static prediction for demands of 20 is observed in all three scenarios. Only in scenario D is an increase in demands of 12 predicted.[31]

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 conducted in an identical manner to the unrestricted design except for two important changes:

- 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

31. This estimation strategy was conducted for values of z_0 in $\left\{\frac{1}{13}, \frac{2}{13}, ..., \frac{12}{13}\right\}$. The direction of the comparative static for demands of 20 does not change with the various assumptions on z_0. Demands of 12 are only predicted to increase significantly with values of z_0 larger than $\frac{10}{13}$. See section 2.4 of the Supplementary Appendix for further details.

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 computer players 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*, the set of announcements was $\{15,18,20\}$; in the second, *R4*, the set was $\{8,15,18,20\}$. In addition, a control treatment for the restricted design, referred to as *R0*, was run to provide a benchmark for observed behavior in *R3* and *R4*. Subjects in *R0* were able to make announcements from the same set as *R3*; however, there were no computer players included.[32] The characteristics of each treatment are summarized in Table 1.

Demands of 15 and 20 were included 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 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. This test is made using the *R3* treatment, where all demands are aggressive. If all subjects correspond to rational types then they should mimic the 18-demand more often than the more aggressive 20-demand. Even allowing for the possibility of behavioral type subjects, an observation of significantly more 20 announcements than 18 announcements can only be rationalized by the model if there are more 20-types in the subject population than 18-types.

In the *R4* treatment, a demand compatible with all other types, including the most demanding type 20, was included. In equilibrium, the model predicts that such demands of 8 should *not* be made by rational types. 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 then their demand. In *R3* 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.[33] For all statistical tests, two p-values are reported. The first comes from a regression-based approach that uses all observations and cluster-robust standard errors that allow for arbitrary correlation between observations from the same session. These are denoted by p. The second comes from an analogous non-parametric test that uses session averages and is denoted by p_n.[34]

32. We would like to thank an anonymous referee for suggesting a control treatment for the restricted design.

33. The first 5 periods are dropped 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. Section 3 of the Supplementary Appendix contains a replication of the main analyses using all 15 periods of the data. For the analysis of second-stage data, only matches where both players are human subjects are included.

34. Section 3 of the Supplementary Appendix contains a detailed description of all the reported statistical tests. For the regression-based approach, the appropriate dependent variable is regressed on treatment indicator variables. For first-stage data, a subject-level random-effect is also included. To conduct inference, a *t*-test or Wald test, using robust standard errors clustered at the session level, is calculated. The clustering is used to account for the fact that subjects within

TABLE 3
Summary of bargaining outcomes

	Treatment					
	C0	*U1*	*U2*	*R0*	*R3*	*R4*
Mean points per period	13.4	13.7	13.7	13.8	14.3	14.4
Percent of second stages	71.4	66.4	77.2	70.0	74.5	73.4
Percent of pie wasted	10.5	8.8	8.8	7.9	4.7	4.1
... conditional on second stage	14.7	13.3	11.4	11.3	6.3	5.5

Table 3 summarizes outcomes from the bargaining process in the six treatments. On average the difference between the points subjects earned across treatments is small, with matches in the unrestricted design wasting more of the pie. In each treatment the majority of matches enter a second stage; for the *U2* treatment and the two restricted design treatments with induced types this rate is larger. 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.

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* relative to *C*0, and occur more frequently in that treatment than announcements of the induced 12 demand do in *U2*.

In the restricted design sessions, the vast majority of demands are for 15 or 20. In *R0*, demands of 15 are most frequent. In *R4*, demands of 20 occur the most, whereas there is a nearly equal rate of demands of 15 and 20 in *R3*. In both of these treatments there are few announcements of 18 (less than 10%), and even fewer announcements of 8 in *R4* (less than 5%).

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 [$p, p_n < 0.01$]. This result is shown in Table 5, which presents the average points payoff of subjects conditional on the announcement made.[35] In the sessions with induced types (treatments *U1*, *U2*, *R3*, and *R4*), announcements of 20 lead to a greater payoff on average than announcements of 15 [$p, p_n < 0.01$]. Making concessionary announcements (*i.e.* for less than half the pie) lead to lower payoffs on average [$p, p_n < 0.01$]. There are no demands that on average lead to more than half the pie.

a session are randomly matched between periods. The non-parametric approach is also reported. For point predictions a two-sided sign-test is used; for matched data, a two-sided signed-rank test, and for comparative statics, a two-sided rank-sum test. Both approaches have advantages and disadvantages — Fréchette (2012) for a discussion of some of these. Reporting both shows that our main results are robust to the statistical approach.

35. Note 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.

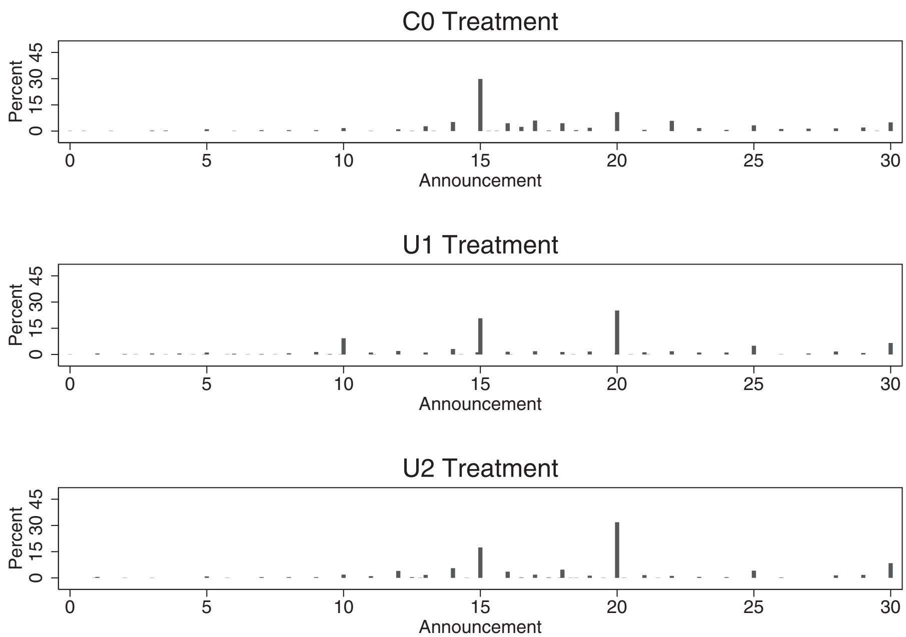


FIGURE 1

Subject announcements in the unrestricted design

TABLE 4

Summary of key announcements

Treatment		Proportion of announcements of 8		10		12		15		18		20	
C0		0.5	(3)	1.7	(10)	1.0	(6)	29.8	(173)	4.5	(26)	10.9	(63)
U1		0.7	(5)	9.3	(65)	2.0	(14)	20.7	(145)	1.4	(10)	25.1	(176)
	p		0.630		0.000		0.215		0.307		0.018		0.044
	p_n		0.572		0.009		0.595		0.347		0.209		0.076
U2		0.4	(3)	1.9	(13)	4.0	(28)	17.4	(122)	4.7	(33)	31.9	(223)
	p		0.819		0.873		0.053		0.152		0.897		0.000
	p_n		0.906		0.916		0.172		0.251		0.675		0.028
R0		.	(.)	.	(.)	.	(.)	55.3	(387)	9.6	(67)	35.1	(246)
R3		.	(.)	.	(.)	.	(.)	49.4	(321)	6.2	(40)	44.5	(289)
	p		.		.		.		0.460		0.117		0.181
	p_n		.		.		.		0.465		0.172		0.347
R4		4.5	(27)	.	(.)	.	(.)	39.0	(234)	9.8	(59)	46.7	(280)
	p		.		.		.		0.005		0.925		0.065
	p_n		.		.		.		0.076		0.917		0.346

Number of observations are given in parentheses.
The p-values give the significance of the difference between the control and the treatment listed in the prior row.

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

TABLE 5
Average payoffs in points

Treatment	Average Payoff to Announcements of 8	10	12	15	18	20
C0	4.2 (2.85)	10.8 (1.03)	12.6 (0.95)	14.0 (0.52)	14.4 (1.02)	13.1 (0.35)
U1	8.2 (0.25)	11.0 (0.50)	11.4 (1.07)	12.6 (0.52)	12.8 (0.63)	13.7 (0.86)
U2	11.0 (1.64)	12.5 (0.68)	13.2 (0.46)	13.3 (0.24)	13.8 (0.36)	14.8 (0.30)
R0				13.9 (0.21)	14.0 (0.44)	13.4 (0.44)
R3				12.8 (0.36)	11.5 (0.55)	14.4 (0.27)
R4	9.9 (0.28)			13.9 (0.28)	13.0 (0.39)	14.5 (0.24)

Weighted average calculated using a random-effects regression of points earned on a complete set of treatment indicators, allowing for a subject-specific effect. Parentheses give robust standard errors, clustered at the session level.

that holds irrespective of the distribution of behavioral types. As a result, the analysis is not dependent on either assumptions over the distribution or on the estimation of this distribution.[36] Table 6 presents summary information on delay in the second stage, in particular comparing the mean delay to this upper bound for subgames with at least 15 observations. The observed delays are significantly longer than their respective upper bounds for all treatments [$p, p_n < 0.01$]. 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 3.17, 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.

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 smaller at 4.70, compared to 6.64 in the unrestricted design. The difference between the designs is particularly notable for treatments with computer players — that is, *U1* and *U2* versus *R3* and *R4* — and is significant [$p, p_n < 0.05$]. However, this difference is statistically insignificant if only explicit subgames are used — that is, subgames involving announcements that might have been made by a computer or announcements of 15 [$p, p_n > 0.1$].[37] Comparing treatments within the restricted design provides further evidence that explicit subgames have lower delays [$p, p_n < 0.01$]. 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.[38]

36. As 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.

37. For *U1*, only the subgames 15–20, and 20–20 are explicit; for *U2*, 12–20, 15–20, and 20–20. For *R3* and *R4* all subgames are explicit by design. For *C0* and *R0*, none of the subgames are explicit since there are no computer players, and an announcement pair of 15–15 does not result in a second–stage.

38. See section 3.3 of the Supplementary Appendix for further evidence of excessive delays.

TABLE 6
Second-stage delay

Treatment	Subgame[§] α_L	α_H	Obs Freq	%	Delay (Seconds) Mean	Bound	Ratio[§§]
C0	*All*		180		248.1		5.80
	15	20	17	9.4	162.5	20.0	
U1	*All*		162		202.7		6.61
	15	20	32	19.8	51.9	20.0	
	20	20	24	14.8	378.2	50.0	
U2	*All*		183		131.8		7.50
	15	20	29	15.8	93.6	20.0	
	20	20	27	14.8	92.2	50.0	
R0	*All*		245		159.3		5.84
	15	18	38	15.5	64.0	11.1	
	15	20	139	56.7	103.4	20.0	
	18	20	27	11.0	209.9	36.4	
	20	20	40	16.3	406.1	50.0	
R3	*All*		193		69.0		4.53
	15	20	119	61.7	53.2	20.0	
	18	20	16	8.3	135.8	36.4	
	20	20	49	25.4	87.0	50.0	
R4	*All*		160		60.9		3.17
	15	18	15	9.4	25.4	11.1	
	15	20	73	45.6	26.1	20.0	
	18	20	17	10.6	93.3	36.4	
	20	20	51	31.9	111.8	50.0	

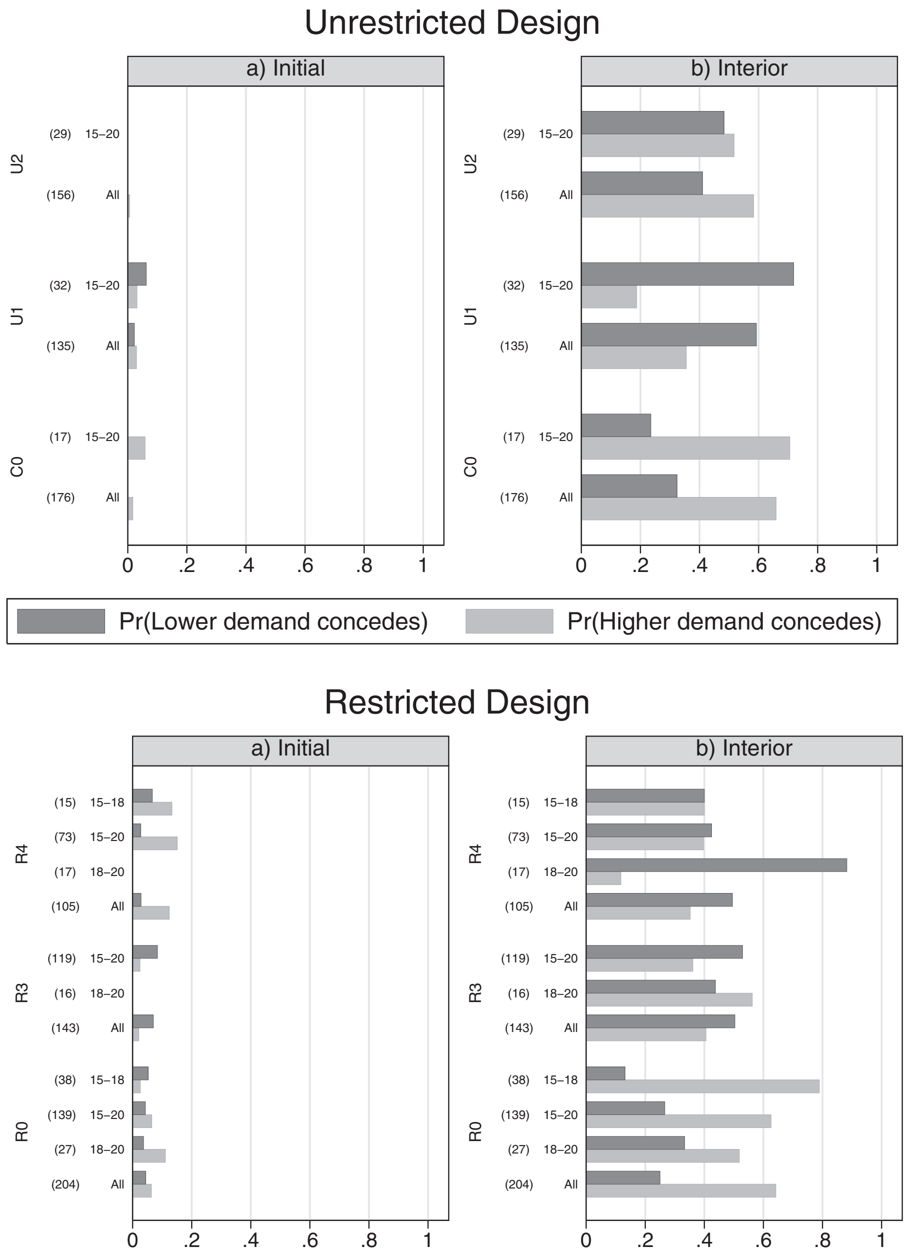
§ Only subgames with at least 15 observations reported.
§§ Weighted (frequency) average of mean delay divided by bound.

5.3. *Concession in the second stage*

The model imposes structure on the pattern of concession behavior. Rational subjects will either always concede instantly, if the demand made by the other player is not mimicked in the first stage, or employ a mixed strategy over the time to concession. 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 15 observations in the unrestricted design.[39] In these sessions, it is not clear that interior concession is in line with the predictions of the model. In the control sessions, the subject making the higher announcement is actually observed to concede more often. While the two treatments with computer players have a significantly larger probability of concession by the lower-announcer than the control [$p, p_n < 0.05$], it is not possible to reject the null hypothesis that the probability is less than or equal to that for the higher-announcer. As with delay, concession behavior in explicit subgames is overall closer to the pattern predicted by the theory than that in other subgames [$p < 0.01, p_n < 0.1$]. With respect to initial concession behavior, there is little

39. Attention is restricted to asymmetric subgames since only these have a clear theoretical prediction. In 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 Supplementary Appendix for further details.



Asymmetric subgames with at least 15 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

evidence that players concede initially in any of the unrestricted design treatments, despite a non-aggressive type being induced in *U2* and a complementary type emerging 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: in *R3*, the lower-announcer is observed conceding more often in both the initial and the 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 in one of the two types of test [$p<0.01, p_n>0.1$].

Finally, in the restricted design it is clear that behavior in explicit subgames is significantly closer to the predicted pattern — that is, behavior in the *R3* and *R4* treatments, where all demands corresponded to an induced type, is much closer to the predicted pattern than is the case in *R0*, where there are no induced types. In particular, interior concession in *R0* is significantly less likely to be by the lower-announcer than is the case in *R3* and *R4* [$p, p_n<0.01$]. Furthermore, it is possible to reject the hypothesis that the lower-announcer is more likely to concede than the higher-announcer [$p<0.01, p_n>0.1$], something that is not possible in any of the treatments with computer players.

6. DISCUSSION

6.1. *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 71.4% 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, Kennan, and Sopher, 1991). However, since incompatible first-stage demands do not imply outright disagreement in our environment, another interesting comparison is the fraction of the pie that is destroyed. We find 10.5% 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.

6.2. *Recognizing the role of reputation*

The first prediction the experimental design sought to investigate was if 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, Cain, and Dawes (2006) and Dana, Weber, and Kuang (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 strategic is a substantive difference, not least since 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.

In their implementations of the ultimatum game with asymmetric information, Mitzkewitz and Nagel (1993) and Güth, Huck, and Ockenfels (1996) find evidence that proposers are inclined to hide behind the possibility of a small pie. That is, when responders do not know the size of the amount to be divided, proposers make offers that would correspond to a fair offer if this amount were smaller than it really is. As with earlier unstructured bargaining studies, these experiments show subjects are willing to use available fairness norms in a strategic manner. While these environments are strategic, the nature of the interaction remains substantively different, as is the interpretation of the mimicking behavior of subjects: the stylized game is symmetric with two-sided incomplete information, and mimicking the demands of the induced types has no possibility of being a fair offer.

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 (2010) also find behavior consistent with reputation building without needing many repetitions.[40] 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.

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

6.3. *Further analysis of announcement strategies*

Several other features of subject behaviour warrant further discussion. First, in the *U1* sessions, not only is there an increase in 20 announcements compared to the control, but also the introduction of two 20 type computers leads to an increase in demands of 10.[41] 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 at least 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 article 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).[42]

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.[43]

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.

6.4. *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. This observation is replicated more precisely using the restricted design, where second-stage behavior is closer to the predictions of the model when the set of behavioral types is explicit — compare *R3* and *R4* with *R0* in Table 6 and Figure 2. In particular, while subjects make demands for more than half of the pie in *R0*, their subsequent concession behavior suggests that such bargaining positions prove to lack credibility since, when they meet a 50–50 demand, they end up waiting longer but being more likely to concede in the end. Taken together, these results suggest that uncertainty over the set of behavioral types is responsible for a significant proportion of the observed departures from the theory, especially in the unrestricted design.

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

42. While 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.

43. In addition, as will be discussed below, the evidence from second-stage behavior in *R0* suggest that, without the presence of a 20-computer player, demanding 20 is not a credible bargaining position.

6.5. *Remaining deviations and extensions of the theory*

While part of the identified deviations can be accounted for by uncertainty over the set of behavioral types, some remain: there is evidence that subjects make 20 announcements too often in *R3* and remain in the concession stage too long in both *R3* and *R4*. It is, however, worth stressing that the quantitative details of the sequential equilibrium predictions are 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.

Overall, prior experiments suggest that subjects might have difficulties implementing such strategies. Although there is quite some evidence from market entry games that aggregate behavior tends to converge to the (mixed-strategy) Nash equilibrium predictions (for example, Erev and Rapoport, 1998; Rapoport, Seale, and Winter, 2000; Duffy and Hopkins, 2005), a number of studies (such as Shachat, 2002) have noted that subjects find mixed strategies difficult to implement at an individual level. 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.[44] Consequently, it is not too surprising to find deviations with respect to these details. However, the systematic nature of those identified are informative, and suggest some possible modifications to the baseline theory.

One extreme assumption of the model is that behavioral types never concede. In line with the model, this is how we programmed our computer players 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, 50–50 types from the subject population might eventually concede results in both greater delay, conditional on eventual agreement, and an increase in more demanding announcements. Thus, a minor modification to the theory brings the finer details of the equilibrium closer to the observed behavior.[45]

This approach to extending the theory focussed on a minimal adjustment to the behavior of the obstinate, “irrational”, types in order to accommodate the observed deviations from the baseline theory; one that reflects the nature of the experimental design. An alternative would be to adjust the preferences of the rational types, for example to incorporate social preferences or fairness goals. A similar analysis can be conducted to see if a small change to preferences might move the predictions of the theory in the direction suggested by the experiment. Assuming that the change does not result in drastically different behavior by rational types in the concession stage, modifications that make concession less appealing for players making *more* demanding announcements, would achieve this goal.[46] In contrast, modifications where the main effect is to make concession less appealing to players making *lower* announcements — for example, envy or spite from those announcing the 50–50 split when they meet unfair opponents — are likely to result in ambiguous predictions for delay and first-stage announcements.

44. Delaying 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.

45. This modification is in the spirit of the perturbation of the standard alternating-offers bargaining given in Binmore and Swierzbinski (2006). See Sections 4.1–4.3 of the Supplementary Appendix for further details.

46. See Section 4.4 of the online appendix for further details. An example of such a modification would be to suppose that players who announce 20 in the first stage suffer a small utility cost of shame if they concede to a fair announcement in the second stage. We thank an anonymous referee for this suggestion.

7. CONCLUSIONS

In implementing the stylized model of Abreu and Gul (2000), the experiments presented in this article 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: there is a 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. 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. We propose a minor modification to the assumed behavior of unresponsive types that would result in predictions with more demanding first-stage announcements and greater delay. Finally, the performance of the model improved markedly in sessions with the restricted design and induced types, where second-stage behavior come closer to the 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 sessions without such explicit types 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 modelled directly.

To summarize, the evidence presented in this article 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.

Acknowledgments. We wish to thank Jim Andreoni, Colin Camerer, John Kagel, Muriel Niederle, Tom Palfrey, Ennio Stacchetti, Leeat Yariv and three anonymous referees, as well as seminar participants at Caltech, Maastricht University, New York University, Paris School of Economics, Royal Holloway, Universitat Autonòma 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, Econometric Society World Congress, the CEA Annual meetings, and the SITE Conference on Experimental Economics for helpful comments and suggestions that improved the article. We thank Raj Advani, Samuel Mati and Anwar Ruff for software development, and Jacopo Perego, Tobias Salz, Chloé Tergiman, Séverine Toussaert and Sevgi Yuksel for research assistance. Fréchette and Lehrer respectively wish to thank the NSF (grants SES-0519045, SES-0721111, SES-0924780, and SES-1225779), CV Starr Center, Center for Experimental Social Science (CESS), and SSHRC for research support. We are responsible for all errors.

Supplementary Data

Supplementary materials are available at *Review of Economic Studies* online.

REFERENCES

ABREU, D. and GUL, F. (2000), "Bargaining and Reputation," *Econometrica*, **68**, 85–117.

Abreu, D. and SETHI, R. (2003), "Evolutionary Stability in a Reputational Model of Bargaining", *Games and Economic Behavior*, **44**, 195–216.

ANDERHUB, V., ENGELMANN, D. and GÜTH W. (2002), "An Experimental Study of the Repeated Trust Game with Incomplete Information", *Journal of Economic Behavior & Organization*, **48**, 197–216.

ANDREONI, J. and BERNHEIM B. D. (2009), "Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects", *Econometrica*, **77**, 1607–1636.

ANDREONI, J. and MILLER, J. H. (1993), "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence", *Economic Journal*, **103**, 570–585.

ATAKAN, A. E. and EKMEKCI, M. (2014), "Bargaining and Reputation in Search Markets", *The Review of Economic Studies*, **81**, 1–29.

BABCOCK, L. and LOEWENSTEIN, G. (1997), "Explaining Bargaining Impasse: The Role of Self-Serving Biases", *The Journal of Economic Perspectives*, **11**, 109–126.

BARON, D. P. (2003), "Private Politics", *Journal of Economics & Management Strategy*, **12**, 31–66.

BINMORE, K. and SWIERZBINSKI, J. (2006), "A Little Behavioralism Can Go a Long Way", *Does Game Theory Work? The Bargaining Challenge*, **8**, 257–276.

BOHNET, I., HARMGART, H. TYRAN, J.-R. et al. (2005), "Learning Trust", *Journal of the European Economic Association*, **3**, 322–329.

BOLTON, G. E. and OCKENFELS A. (2000), "ERC: A Theory of Equity, Reciprocity, and Competition", *The American Economic Review*, **90**, 166–193.

BRANDTS, J. and FIGUERAS, N. (2003), "An Exploration of Reputation Formation in Experimental Games", *Journal of Economic Behavior & Organization*, **50**, 89–115.

BULOW, J. and KLEMPERER, P. (1999), "The Generalized War of Attrition", *The American Economic Review*, **89**, 175–189.

CAMERER, C. and WEIGELT, K. (1988), "Experimental Tests of a Sequential Equilibrium Reputation Model", *Econometrica*, **56**, 1–36.

CHARNESS, G. and RABIN, M. (2002), "Understanding Social Preferences with Simple Tests", *The Quarterly Journal of Economics*, **117**, 817–869.

CHATTERJEE, K. and SAMUELSON, L. (1987), "Bargaining with Two-sided Incomplete Information: An Infinite Horizon Model with Alternating Offers", *The Review of Economic Studies*, **54**, 175–192.

CHATTERJEE, K. and SAMUELSON, L. (1988), "Bargaining under Two-Sided Incomplete Information: The Unrestricted Offers Case", *Operations Research*, **36**, 605–618.

COOPER, D. C. and KAGEL, J. H. (forthcoming), "Other Regarding Preferences: A Survey of Experimental Results", *Handbook of Experimental Economics*, **2**.

CRAWFORD, V. P. (1982), "A Theory of Disagreement in Bargaining", *Econometrica*, **50**, 607–637.

DAMIANO, E., LI, H. and SUEN, W. (2012), "Optimal Deadlines for Agreements", *Theoretical Economics*, **7**, 357–393.

DANA, J., CAIN, D. M. and DAWES R. M. (2006), "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.

DANA, J., WEBER, R. and KUANG, J. (2007), "Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness", *Economic Theory*, **33**, 67–80.

DUFFY, J. and HOPKINS, E. (2005), "Learning, Information, and Sorting in Market Entry Games: Theory and Evidence", *Games and Economic behavior*, **51**, 31–62.

ENGLE-WARNICK, J. and SLONIM, R. L. (2004), "The Evolution of Strategies in a Repeated Trust Game", *Journal of Economic Behavior & Organization*, **55**, 553–573.

EREV, I. and RAPOPORT, A. (1998), "Coordination, "Magic," and Reinforcement Learning in a Market Entry Game", *Games and Economic Behavior*, **23**, 146–175.

FALK, A., FEHR, E. and FISCHBACHER, U. (2003), "On the Nature of Fair Behavior", *Economic Inquiry*, **41**, 20–26.

FEHR, E. and SCHMIDT, K. M. (1999), "A Theory Of Fairness, Competition, and Cooperation", *Quarterly Journal of Economics*, **114**, 817–868.

FORSYTHE, R., KENNAN, J. and B. SOPHER (1991), "An Experimental Analysis of Strikes in Bargaining Games with One-Sided Private Information", *The American Economic Review*, **81**, 253–278.

FRÉCHETTE, G. R. (2012), "Session-effects in the laboratory", *Experimental Economics*, **15**, 485–498.

GROSSKOPF, B. and SARIN, R. (2010), "Is Reputation Good or Bad? An Experiment", *The American Economic Review*, **100**, 2187–2204.
GÜTH, W., HUCK, S. and OCKENFELS, P. (1996), "Two-level Ultimatum Bargaining with Incomplete Information: An Experimental Study", *The Economic Journal*, **106**, 593–604.
HÖRISCH, H. and KIRCHKAMP, O. (2010), "Less Fighting than Expected - Experiments with Wars of Attrition and All-Pay Auctions", *Public Choice*, **144**, 347–367.
JUNG, Y. J., KAGEL, J. H. and LEVIN, D. (1994), "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.
KAMBE, S. (1999), "Bargaining with Imperfect Commitment", *Games and Economic Behavior*, **28**, 217–237.
KWIEK, M. (2011), "Reputation and Cooperation in the Repeated Second-Price Acutions", *Journal of the European Economic Association*, **9**, 982–1001.
MCKELVEY, R. and PALFREY, T. (1995), *The holdout game: an experimental study of an infinitely repeated game with two-sided incomplete information*, In W. Barnett, H. Moulin, M. Salles and N. Schofield, (Eds.) *"Social Choice, Welfare, and Ethics"*. (Cambridge University Press, Cambridge).
MITZKEWITZ, M. and NAGEL, R. (1993), "Experimental Results on Ultimatum Games with Incomplete Information", *International Journal of Game Theory*, **22**, 171–198.
MUTHOO, A. (1999), *Bargaining Theory with Applications*. (New York, NY: Cambridge University Press).
MYERSON , R. B. (1991), *Game Theory: Analysis of Conflict*. (Cambridge, MA: Harvard University Press).
NERAL, J., and OCHS, J. (1992), "The Sequential Equilibrium Theory of Reputation Building: A Further Test", *Econometrica*, **60**, 1151–1169.
OCHS, J. and ROTH, A. E. (1989), "An Experimental Study of Sequential Bargaining", *The American Economic Review*, **79**.
ÖZYURT, S. (2011), "Searching a Bargain: Power of Strategic Commitment", Mimeo, Sabanci University.
POVEL, P. (2005), "Multiple Banking as a Commitment Not to rescue", *Research in Finance*, **21**, 175–199.
RAPOPORT, A., SEALE, D. A. and WINTER, E. (2000), "An Experimental Study of Coordination and Learning in Iterated Two-Market Entry Games", *Economic theory*, **16**, 661–687.
ROTH, A. E. (1995), "Bargaining Experiments", *Handbook of Experimental Economics*, 253–348.
ROTH, A. E. and MALOUF, M. W. K. (1979), "Game-Theoretic Models and the Role of Information in Bargaining", *Psychological Review*, **86**, 574–594.
ROTH, A. E. and MURNIGHAN, J. K. (1982), "The Role of Information in Bargaining: An Experimental Study", *Econometrica*, **50**, 1123–1142.
ROTH, A. E. and SCHOUMAKER, F. (1983), "Expectations and Reputations in Bargaining: An Experimental Study", *The American Economic Review*, **73**, 362–372.
ROTH, A. E., MALOUF, M. W. K. and MURNIGHAN, J. K. (1981), "Sociological Versus Strategic Factors in Bargaining", *Journal of Economic Behavior & Organization*, **2**, 153–177.
RUBINSTEIN, A. (1982), "Perfect Equilibrium in a Bargaining Model", *Econometrica*, **50**, 97–109.
SCHELLING, T. C. (1960), *The Strategy of Conflict*, (Cambridge, Mass: Harvard University Press).
SHACHAT, J. M. (2002), "Mixed Strategy Play and the Minimax Hypothesis", *Journal of Economic Theory*, **104**, 189–226.
WILSON, R. K. and ECKEL, C. C. (2006), "Judging a Book by its Cover: Beauty and Expectations in the Trust Game", *Political Research Quarterly*, **59**, 189–202.
WOLITZKY, A. (2012), "Reputational bargaining with Minimal Knowledge of Rationality", *Econometrica*, **80**, 2047–2087.