



The power of outside options in the presence of obstinate types [☆]



Tom-Reiel Heggedal, Leif Helland, Magnus Våge Knutsen ^{*}

Department of Economics and CESAR, BI Norwegian Business School, Norway

ARTICLE INFO

Article history:

Received 18 June 2021

Available online 24 October 2022

JEL classification:

C78

C91

D82

Keywords:

Bargaining

Reputation

Obstinate types

Experiment

ABSTRACT

We experimentally investigate the role of two-sided reputation-building in dynamic bargaining. In the absence of outside options, rational bargainers have an incentive to imitate obstinate types that are committed to an aggressive demand, inducing delay. Outside options remove this incentive and ensure immediate agreement whenever two rational bargainers match. Our data support the hypothesis that outside options cut down on imitation and ensure timely agreements, but only if subjects share a belief about what constitutes obstinacy. Further, we find that outside options are exercised excessively and that efficiency is no better than it is in their absence. We ascribe this result to the presence of fairness preferences in the subject pool.

© 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

It is widely recognized that rational bargainers may profitably exploit a reputation for irrationality. In particular, a rational player may act to create the impression that he is incapable of yielding, hoping this impression will make his opponent believe he has “the last clear chance” and must give in.¹ Irrationality takes many forms.² Formal bargaining theory has largely focused on “obstinacy”—a simple and analytically tractable form of irrationality—in which a bargainer insists on a particular claim and rejects any offer below it.^{3,4} In dynamic bargaining environments, rational bargainers have incentives

[☆] Leif Helland passed away on December 20, 2021. He is greatly missed. We appreciate comments from editor Gary Charness, an anonymous advisory editor, two anonymous referees, Kjell Arne Brekke, Mehmet Ekmekci, Tore Ellingsen, Guillaume Fréchet, Jack Fanning, Topi Miettinen, Espen R. Moen, Ernesto Reuben, participants at the ESA European Meeting, 2017, the ESA Asia Pacific Meeting, 2018, the NCBEE, 2018, the BEET Workshop, 2020, and EEA, 2020. The research is financed by the Research Council of Norway, grant #250506.

^{*} Corresponding author.

E-mail address: magnus.v.knutsen@bi.no (M. Våge Knutsen).

¹ A standard illustration is Richard Nixon's madman theory (see Sagan and Suri (2003) for a thorough discussion).

² The negotiation literature reflects this and is ripe with psychological advice on how to deal with various forms of irrationality (e.g., Raiffa et al. (2002), ch. 16).

³ For imitation of obstinate types to work, such types need to exist in the population at large. Abreu and Sethi (2003) show that obstinate types are not rooted out by evolutionary pressures if rational players incur some (arbitrarily low) cost of implementing their strategy, and if there are compatible pairs of obstinate types in the initial population. A common observation in dynamic bargaining experiments with discounting is that a fraction of subjects make disadvantageous counter offers (e.g., Binmore et al. (1989), Ochs and Roth (1989)). As noted by Embrey et al. (2015), this is what should be expected with obstinate types in the subject sample.

⁴ Ultimatum experiments offer indicative evidence of the existence of obstinate types. Given responder behavior, demands above 80 percent of the pie face rejection rates that fail to maximize proposers' payoffs in these experiments. Still, a non-negligible fraction of such demands is typically observed even in later rounds of these experiments, after learning has taken place and responder behavior is presumably known by proposers.

to imitate the behavior of such types. Without access to outside options, imitation leads to delay and erosion of efficiency. Access to such options may, on its side, provide rational bargainers with a potent tool capable of securing timely agreements and restoring efficiency through deterrence of imitation.

We experimentally investigate the power of outside options in an alternating offer framework with obstinate types. In the absence of outside options, Abreu and Gul (2000) show that if both parties have a positive probability of being an obstinate type who demands a fixed share and never accepts anything less, rational players imitate the behavior of the obstinate types.⁵ As a result, equilibrium takes the form of a war of attrition which in turn leads to delayed agreement and inefficiency. Compte and Jehiel (2002) show that access to outside options can restore efficiency by removing the incentive to imitate obstinate types. If the outside option is higher than the obstinate offer, a rational player will never concede to such an obstinate demand. As such, there is no incentive for rational players to imitate these obstinate types, and agreement between two rational players is immediate.

In most real-life bargaining situations, unilateral termination is permissible, in which case the bargainers obtain some (more or less valuable) outside option. Together with the predictions of Compte and Jehiel (2002), this implies that delay and inefficiencies might be less prevalent than suggested by the framework of Abreu and Gul (2000) and the corresponding laboratory findings of Embrey et al. (2015), both of which ignore outside options. While considerable effort is invested in extending and generalizing the Abreu and Gul (2000) framework (see below), its robustness to low-valued outside options remains an open empirical question. We conduct a first systematic investigation of this pressing question.

The equilibria of the model hinge critically on the distribution of obstinates being common knowledge among rational bargainers. Measuring beliefs about obstinacy in naturally occurring bargaining data poses major challenges, foremost because one's own type is a piece of private information with strategic value. Moreover, assuming shared beliefs about the distribution of obstinacy is a strong assumption in any real-life bargaining situation. However, in a controlled laboratory setting, obstinate bargainers can be induced and their distribution made public knowledge. This creates a degree of control over beliefs and allows us to test the core predictions of Compte and Jehiel (2002) in a direct way.

Our experiment contains four main treatments that vary along two dimensions: Whether outside options are available or not, and whether robot players with an inflexible demand (induced obstinate types) are present or not. Our contribution is threefold.

First, we uncover a large and statistically strong treatment effect of introducing outside options in the presence of obstinate robot players. To break it down: In the absence of outside options, we provide further evidence that subjects recognize the role of reputation in bargaining, corroborating the results by Embrey et al. (2015). Furthermore, and in line with the model, providing subjects with outside options that are not binding in the standard model sharply reduces both imitation and delay.

Second, our results indicate that the power of outside options is contingent on subjects having shared beliefs about obstinacy. In treatments without robot players, reputation-building is possible to the extent that homegrown obstinates are found in the subject pool. We uncover a substantial share of homegrown obstinates in our sample. Rational bargainers have incentives to imitate these when outside options are unavailable. However, we find no treatment effect of outside options on imitation in the absence of robot players. Thus, imitation mainly takes place when there are no outside options *and* there is a shared belief about obstinacy induced by the experimenter. In this way, our experiment accentuates the challenges in testing the model predictions when beliefs cannot be exogenously manipulated (e.g., in naturally occurring bargaining data).

Finally, we find that outside options cause a significant decrease in delay even in treatments with no robot players. Consistent with the model, we observe that subjects opt out as a response to aggressive demands. On top of this, we observe that outside options are used in excess of equilibrium predictions both in the presence and in the absence of robot players. In particular, we find that some subjects are willing to forgo material payoffs by opting out rather than accepting inequitable proposals. We argue that fairness preferences go a long way in rationalizing such responses.⁶ Perhaps surprisingly, we do not find a robust learning effect that washes out aggressive proposals over time. That is, while the share of aggressive demands is higher in early games, it remains high and significant also in later games. We interpret this persistence as signifying a presence of stubbornly aggressive (i.e., homegrown obstinate) subjects in our sample. In sum, the excessive use of outside options offsets their potentially efficiency-promoting effects on imitation and delay in the experiment.

We also offer results from three additional treatments that are designed to further explore the role of robots in shaping beliefs about obstinacy. We do this, first, by severely restricting the strategy sets of subjects, aiding subjects further in coordinating on a shared belief by admitting only one way of obstinacy. Second, we reveal to the bargainers whether the current match was a robot or a human after the termination of a bargaining game. This introduces a minimal social-identity cost of imitation. We find that these twists add little to our main robot treatment without outside options in terms of imitation.

The presence of delay in bargaining has motivated an extensive theoretical literature. One strand focuses on incomplete information as a source of delay in bargaining. The core insights are that incomplete information creates inefficiencies

⁵ To economize on language, we understand “rational” to mean rational and purely self-regarding. Whenever players are assumed to be rational and other-regarding, this is explicitly stated.

⁶ We have chosen not to view subjects that insist on an equal division as “moderate obstinates.” Instead, we find it fruitful to model them as maximizers that incur a utility loss from unfavorable inequality. A “moderate obstinate” would not exercise his outside option, because doing so would leave it with less than half the pie. Consequently, the presence of such types cannot account for our data.

(Myerson, 1979) and produces multiple equilibria.⁷ A different strand of literature, starting with Schelling (1956), suggests that delay is caused by parties attempting to convince one another that they are committed to a particular outcome.⁸ A third strand is represented by the influential contribution of Abreu and Gul (2000).⁹ It highlights the role of reputation in bargaining and shows that delay can arise as an equilibrium outcome when rational bargainers have a common prior belief on obstinacy. Several papers extend this framework by exploring the roles of deadlines (Fanning (2016)), markets with search frictions (Atakan and Ekmekci (2013), Özyurt (2015b), Özyurt (2015a)), the presence of complex behavioral types (Abreu and Pearce (2007)),¹⁰ the influencing of third parties (Perlroth, 2020), and the strategic use of courts (Ekmekci and Zhang, 2020). Models in this strand frequently offer unique equilibrium predictions, providing clear-cut benchmarks for bargaining experiments.¹¹

The experimental literature on two-sided reputation formation in bargaining is slim. To the best of our knowledge, the only existing experiment is found in Embrey et al. (2015). Their set-up follows Abreu and Gul (2000). Bargainers play one round of a Nash demand game. If claims are incompatible, bargainers enter a war of attrition and make repeated binary choices between insisting on their initial claim or giving in. Our set-up is richer. Bargaining takes place in a protocol where claims may be adjusted over time. In addition, outside options are made a part of the bargaining environment.

A substantial experimental literature investigates bargaining with one-sided type uncertainty. The focus in this literature differs from ours. The primary interest is in screening rather than reputation-building (e.g., Bochet and Siegenthaler (2018), with references).¹² A voluminous experimental literature addresses one-sided reputation-building in a small set of generic games—trust games, prisoner’s dilemma games, and centipede games (e.g., Anderhub et al. (2002) and Grosskopf and Sarin (2010), with references).¹³

The article is organized as follows: In section 2 we outline the theoretical framework, in section 3 we present our experimental design and procedures, in section 4 we report the results from our main experiment, and in section 5 we describe and provide the results from our additional treatments. In section 6 we offer a brief conclusion.

2. Theoretical framework

In this section, we provide a brief overview of the bargaining environment as well as the core predictions of the model.¹⁴

2.1. Bargaining without outside options

Two risk-neutral players, $i = 1, 2$, bargain on the division of a pie with a value normalized to 1. The horizon is infinite and the players alternate between being the proposer or the responder of offers until bargaining is resolved. The players are impatient and discount future payoffs by a common discount factor $\delta < 1$. A player’s payoff is equal to the share of the pie he obtains.

Ex ante, the players are obstinate with some probability. Obstinate players are characterized by an inflexible demand and reject any offer below this demand. Define $S := \{\theta_1, \dots, \theta_M\} \subset [\frac{1}{2}, 1]$ as the set of obstinate types with respective demands θ_m , with $\theta_{m+1} > \theta_m$. This implies that we allow for types who demand less than v^* , where $v^* = \frac{1}{1+\delta}$ denotes the equilibrium payoff of the proposer in a game without obstinate types. We assume that $\theta_m > v^*$ for at least one type. The probability that a player is a θ_m -type is ε_m , for $m = 1, \dots, M$. The probability that a player is rational is denoted $\varepsilon_R = 1 - \sum_{m=1}^M \varepsilon_m$. While the structure of the game is common knowledge, a player’s own type is assumed to be privately known.

The key equilibrium property is that rational players have incentives to make demands that mimic those of obstinate types. To understand this, consider an equilibrium candidate in which rational players do not mimic, but demand v^* and accept $1 - v^*$. If rational players demand only v^* and always accept $1 - v^*$, any offer of θ_m (for $m = 1, \dots, M$) or a rejection of $1 - v^*$ must come from an obstinate player. Thus, when faced with such behavior, a rational player must infer that he

⁷ For multiplicity of equilibria in the case of one-sided private information see (Rubinstein, 1985), and for two-sided private information see (Fudenberg and Tirole, 1983). Several routes have been explored in order to achieve uniqueness, among them, severely restricting the strategy sets (Chatterjee and Samuelson, 1988), introducing contestable selection criteria (Sobel and Takahashi, 1983; Chatterjee and Samuelson, 1987; Cramton, 1984), and imposing axiomatic restrictions on equilibrium (Rubinstein, 1985; Gul and Sonnenschein, 1988).

⁸ Formalization of these complete information models is found in Ellingsen and Miettinen (2008), Ellingsen and Miettinen (2014), and Crawford (1982).

⁹ Abreu and Gul (2000) build on the framework of Kreps and Wilson (1982) and Milgrom and Roberts (1982). An early application of one-sided reputation-building in bargaining is found in the r -insistent types of Myerson (1991), pp. 399–403.

¹⁰ Complex behavioral types play time paths with increasingly or decreasingly aggressive demands.

¹¹ The literature also offers other explanations for delay. Schweighofer-Kodritsch (2018) show that heterogeneity in time preferences can cause delay in the equilibrium of the alternating-offer model even under complete information and common knowledge of rationality. Kim et al. (2021) conduct an experimental test of this model, and find support for its qualitative predictions. Using epistemic game theory Friedenber (2019) shows that lack of common knowledge rationality in alternating-offer bargaining may in itself produce delayed agreement as an equilibrium phenomenon. On the less formal side, self-serving biases have been shown to cause impasse and delay in unstructured bargaining experiments (Babcock and Loewenstein (1997)), while Brekke et al. (2022) uncover delay in experiments when parties form reference points on bargaining outcomes.

¹² Typically, an uninformed seller faces a distribution of buyers that vary in their valuations and is given the opportunity to revise prices over time. The conditions under which the Coase conjecture (Coase (1972)) holds are at the center. Fanning and Kloosterman (2022) find strong experimental evidence for a variant of the Coase conjecture conditioned on subjects having private information on their own fairness preferences.

¹³ A recent focus in this strand of literature has been on the interaction between reputation and competition (Huck et al. (2016), with references).

¹⁴ See online appendix A for the formal analysis.

is facing an obstinate player. However, given these beliefs, a rational player could deviate and demand and obtain θ_m in the first round if his opponent is rational. Consequently, there cannot be an equilibrium without imitation. Moreover, the most aggressive type must be mimicked with positive probability, and offers equal to $1 - \theta_M$ are rejected with a positive probability, which in turn leads to an expected delay in resolving bargaining (see Proposition 1 in online appendix A).

2.1.1. Bargaining with outside options

Turning to outside options, we consider an environment in which players have the opportunity to opt out whenever it is their time to make a decision. If a player opts out, both players receive their outside options, denoted by v^{out} . The outside option value is restricted such that $v^{out} < \varepsilon_R v^* + (1 - \varepsilon_R)\delta v^{out}$. This implies that the outside option is non-binding in the sense that it would never be used in the case where $\varepsilon_R = 1$. It also implies that a proposer will always make a proposal in the first round.

Assume that there is an obstinate type θ_K such that $1 - \theta_K = v^{out}$, where $1 \leq K < M$.¹⁵ This cut-off ensures that the outside option is (weakly) preferred by a rational player to the offer from an obstinate type $\theta_m \geq \theta_K$. Then there is no incentive to imitate any type $\theta_m \geq \theta_K$ in equilibrium.¹⁶ To understand this, notice that in equilibrium, players do not believe rational players demand more than θ_K . Thus, faced with a demand larger than θ_K , a player's updated belief is that he is facing an obstinate type and he is better off taking his outside option. Then, the most a rational player can obtain by imitation of an obstinate type $\theta_m \geq \theta_K$ is v^{out} (in case he meets another rational player), and he is indeed better off making an initial demand lower than θ_K .

If ε_m is positive in the range $\{\varepsilon_1, \dots, \varepsilon_{K-1}\}$, then players have incentives to make demands that mimic types in this range because outside options remove only incentives to imitate types larger than θ_K . That is, imitation must still take place in equilibrium as long as $\theta_m > v^*$ for at least one type $m < K$, and there will still be delay in the resolution of bargaining (see Proposition 1 with discussion in online appendix A). On the other hand, if ε_m is positive only in the range $\{\varepsilon_K, \dots, \varepsilon_M\}$, equilibrium predictions on the effect of outside options on delay are starker. In this case, rational players will not imitate obstinacy. Rather, a pair of rational players will come to an agreement in the first round and agree on the division $(v^*, 1 - v^*)$. Moreover, still in this latter case, in bargaining between a rational player and an obstinate player, bargaining is terminated in the first or the second round, depending on the type of player making the initial offer. The payoff to the rational player is v^{out} if the obstinate type makes the initial offer, and δv^{out} if the rational type makes the initial offer.

In section 4.2, we consider a richer environment in which we allow for optimizing agents with an aversion to unfavorable inequalities. We show that in the presence of such aversion, a given outside option will deter imitation of a wider range of obstinate types than if all optimizing agents were purely self-serving.

3. Treatments, design, and procedures

In our experiment, we implement an alternating-offer bargaining protocol where subjects bargain over a shrinking pie.¹⁷ The four main treatments vary with respect to induced obstinacy and the availability of outside options.

Our first two treatments use robots to induce obstinate demands: In *RN* (**R**obot / **N**o outside option) outside options are absent while in *RO* (**R**obot / **O**utside option) they are present. Our next two treatments are without robots: In *NN* (**N**o robot / **N**o outside option) outside options are absent while in *NO* (**N**o robot / **O**utside option) they are present.

3.1. Design

The aim of our design is to investigate how outside options influence imitation behavior. In particular, we focus on aggressive demands. Two features of the framing of our experiment help facilitate imitation of such demands. First, obstinate types are induced by robot players and the distribution of such types is made public knowledge. The robots in our experiment demand 70 percent of the pie. Second, across all treatments there is a focal point on making a demand equal to 70 percent of the pie. The idea of the focal point is to set up the environment in favor of imitating an obstinate demand. Further, to create some observational distance between the robot demand and v^* in the data, we induce a common discount factor by letting the pie over which the subjects bargain shrink by a fixed rate of 10 percent after each round. This gives $v^* = 0.53$.¹⁸

The theory predicts that subjects will imitate robot demands in the absence of outside options. This imitation will in turn lead to delay. The value of the outside option in our experiment is 1/3 of the pie, which is strictly higher than the obstinate offer. Consequently, in treatments with outside options, we expect to observe no imitation of robot demands.

¹⁵ This assumption simplifies the exposition and is without loss of generality.

¹⁶ This follows from Compte and Jehiel (2002). Note that Compte and Jehiel (2002) consider an environment where parties have only one obstinate type. However, their result extends to a multiple-type environment; see footnote 11 in Compte and Jehiel (2002).

¹⁷ The online appendix G contains the full set of instructions used in the experiment.

¹⁸ A further reason to induce a v^* close to 0.5 is that the empirical evidence on how forcefully discounting can move demands away from 50-50 is mixed. See Kim et al. (2021) for an excellent overview of this topic.

Our primary outcome measures are initial demands in a given game of bargaining, and delay. We focus on initial demands, rather than on demands made in all periods, for the following reasons: In theory, bargaining games that go beyond the first period in treatment *RN* will consist of both rational players who are imitating obstinacy and obstinate types. In treatment *RO*, theory predicts that rational players should not imitate obstinate types with demands larger than $2/3$ (such as the robots), and we expect bargaining games that go beyond the first period to consist of pairs with at least one obstinate bargainer. Consequently, while games in *RO* that go beyond the first period can have a high share of obstinate demands, this may not be due to imitation, but rather to the presence of obstinate types. Including demands from all periods therefore risks masking the effect we are interested in, i.e., the effect of outside options on the degree of imitation.

The treatments *RN* and *RO* allow for tests of the main predictions from Compte and Jehiel (2002); outside options reduce both imitation behavior and delay. Imitation of robot players is not directly observed in the data but is inferred from demands that are equal to the robot demand ($x = 0.7$). Thus, the difference in the share of robot demands between *RN* and *RO* returns a measure of difference in the imitation of robots.

It is not reasonable to postulate that all subjects in the sample are fully rational. Under the assumptions of the model, demands of $x > 2/3$ are strictly irrational in treatments *RO* and *NO*. That is, as long as subjects have shared beliefs about types that inflexibly insist on $x > 2/3$, a rational player facing such a demand will use his outside option immediately. Consequently, nothing can be gained from making such a demand. In the following, we refer to subjects making demands $x > 2/3$ in treatments with outside options as “homegrown obstinates,” and refer to any demand $x > 2/3$ as “an obstinate demand.” Through the lens of the model, then, the share of demands $x > 2/3$ in *RO* provides an estimate of homegrown obstinates in the sample.¹⁹ Note that, due to randomization, the expected distribution of homegrown obstinates will be the same across treatments. Thus, the difference in the share of obstinate demands between *RN* and *RO* provides a direct measure of imitation of obstinacy in *RN*, also when there are homegrown obstinates in the population.

NN and *NO* allow for exploring the role of shared beliefs about obstinacy in our subject sample. There are no robot players in these treatments to induce obstinate types, but if homegrown obstinates with demands $x > 2/3$ are believed to be present in the subject pool, theory predicts that imitation of such types should take place in *NN*. As above, taking the model at face value, the share of demands $x > 2/3$ in *NO* provides us with an estimate of homegrown obstinates in our sample, while the difference between *NN* and *NO* provides a measure of imitation of such homegrown obstinates in *NN*. Finally, the difference in the share of obstinate demands between *RN* and *NN* gives the effect of induced obstinates on imitation.

3.1.1. Homegrown 50-50 types

Clearly, there can also be “moderate obstinate” subjects in the pool that insist on an equal split of the pie or some other share below $2/3$, as well as rational subjects that have incentives to imitate such types. The most natural homegrown type is perhaps 50-50 types, and to understand how such types may affect theoretical predictions for our treatments, we discuss two stylized scenarios. First, consider a scenario where the only homegrown types are 50-50 types, and that this is a shared belief among rational players. In treatment *RN*, the robot demand (0.7) is mimicked with positive probability, and there is an incentive to mimic the 50-50 types as well when players are sufficiently patient (see Proposition 2 parts 2 and 3 in online appendix A). In treatment *RO*, however, there is no incentive to mimic the 50-50 types nor the robot demand. The outside options remove imitation of robot demands, and thereby the potential gain of being perceived as a 50-50 type. Because $v^* > 0.5$, rational proposers are then better off demanding v^* in the first round, as long as the probability of facing a 50-50 type is not too large and the player is not too impatient.²⁰ Moreover, since $1 - v^* > \delta 0.5$, there is nothing to gain for a rational responder by rejecting the equilibrium offer and then imitating the 50-50 type in the next round. Similarly, in treatments *NN* and *NO*, there is no incentive to imitate 50-50 types since there is no imitation of other behavioral types when the probability of players being any other behavioral type than 50-50 is zero (see Proposition 2 part 1 in online appendix A).

Second, consider a scenario with both 50-50 types and obstinate 0.7 types (the robot demand) as homegrown, and that this scenario is a shared belief among the rational players. In this scenario, there is a positive probability that a player faces either of the behavioral types across all treatments. Thus, in the treatments without outside options, *RN* and *NN*, the obstinate 0.7 types are mimicked with positive probability, and there are incentives to mimic the 50-50 types as well when players are sufficiently patient. In contrast, in treatments *RO* and *NO*, there are no incentives to imitate the obstinate 0.7 types and thus neither the 50-50 types.

The upshot is that the presence of 50-50 types does not affect under which conditions we expect to observe imitation of robot demands. Hence, while the presence of 50-50 types may affect how we interpret level effects in our experiment, it does not affect our main results qualitatively, because our experiment is designed to tease out the imitation of types that demand more than $2/3$ of the pie. Given that the expected distribution of 50-50 types is the same across treatments, the difference in obstinate demands between *RN* and *RO* provides a lower bound on the share of subjects making demands consistent with rational equilibrium play. We calculate this bound as the difference in the share of subjects with demands $x > 2/3$ between treatments *RN* and *RO*. However, a similar calculation is not meaningful for demands $x = 0.5$, since we in

¹⁹ This estimate of obstinacy includes subjects that irrationally choose the robot demand ($x = 0.7$) due to, e.g., a focal point effect.

²⁰ Potentially facing 50-50 types, the condition for a rational player to not immediately yield when having the first-round proposal is $v^* \geq (1 - \delta)0.5 + \frac{\delta 50}{\epsilon_R} 0.5$.

Table 1
Treatments, matching blocks, and subjects.

Treatments	Number of subjects	Blocks	Games per block (mean)
RN (Robot/No option)	96	12	9.2
RO (Robot/Option)	96	12	9.7
NN (No robot/No option)	40	5	10
NO (No robot/Option)	40	5	10

the data cannot separate obstinate 50-50 demands from rational v^* demands because v^* is close 0.5. Thus, we cannot infer much about rational equilibrium play from the difference in (approximately) 0.5 demands between *RN* and *RO*, since this difference is driven by both lower (no) imitation of 50-50 types and more play of v^* in *RN* by rational players.

3.2. Procedures

In the experiment, the roles (proposer and responder) alternate until an agreement is reached, until one of the players opts out (in treatments with outside options), or until the value of the pie falls below a certain threshold (1 krone which equals approximately 1 percent of the original pie).²¹ Demands, offers, outside options, and payoffs are denominated in experimental currency units (ECU). The exchange rate is set to equalize expected payoffs between treatments. We induce obstinacy by including pre-programmed robot players with a fixed and inflexible demand corresponding to a 0.7 share of the pie. The robot demand and the outside options are discounted at the same rate as the pie.²²

When subjects formulate demands, they can either click on a button corresponding to a demand fixed at 0.7 of the pie, or move to the next screen and demand any share $x \in [0.00, 1.00]$ (or $x \in \{0.5, 0.7\}$, see below). The “button feature” is the same in all treatments in order to hold focal point effects constant.²³

We use matching blocks of 8 subjects and 2 robots in robot treatments, and 8 subjects in no-robot treatments.²⁴ Subjects stay within blocks, and unique subjects are used in all treatments. In the analysis, we regard average subject behavior within blocks as independent observations.²⁵ Subjects are randomly re-matched (within blocks) over games, and play up to 10 games. Robots are not paired with other robots, so the probability of a subject being matched with a robot is 25 percent. In each bargaining pair, roles as first and second movers are randomly assigned prior to each game. Due to time constraints and heterogeneity in decision times, the number of games played varies over sessions, with a maximum of 10 games and a minimum of 7 games. On average, fewer games were played in treatments without outside options; see *Games per block (mean)* in Table 1.

Sessions were conducted in the Research Lab at BI Norwegian Business School in Oslo, and at the LEE Lab at the University of Copenhagen during the period June 2017 to October 2019. Subjects were recruited from the general student populations of BI Norwegian Business School, the University of Oslo, and the University of Copenhagen, respectively. Recruitment and subject management were administered through ORSEE (Greiner, 2015). A total of 392 subjects participated. In total, 2180 bargaining games were played. On average, subjects in the Norwegian sessions earned 24 EUR while subjects in the Danish sessions earned 32 EUR. The protocol was implemented in zTree (Fischbacher, 2007).

A pre-study plan for the experiment was posted at the AEA RCT registry on January 17, 2018 (after data from the pilot were collected).²⁶ The plan covers our four main treatments. For these treatments, we report in accordance with the plan. Based on a pilot included in the pre-study plan, a power of more than 90 percent for the treatment effect on the difference in the share of obstinate demands ($RN - RO$) was calculated to require a total of 24 matching blocks in a balanced design (given a 5 percent significance level and a Wilcoxon rank test).²⁷

4. Results

The results that follow exclude observations from the first three games to allow for learning.²⁸ We primarily focus on two outcomes from the experiment: first-round demands and delay. We report two different p -values. First, p -values from

²¹ Data were collected in both Norway and Denmark (see below). One krone is the smallest coinage in both countries. At the time of the experiments, one DKK (NOK) was worth 0.14 (0.10) EUR.

²² See Heggedal and McKay (2022) for a discussion of discounting procedures in experiments.

²³ One may ask why we do not follow Embrey et al. (2015) in using a lottery procedure, rather than using linear monetary payoffs. In theory, the lottery procedure will eliminate risk attitudes (Roth and Malouf, 1979). Whether it succeeds at this in practice, however, is unclear (Berg et al., 2008). We decided to stick with linear monetary payoffs to make the experiment more transparent to subjects.

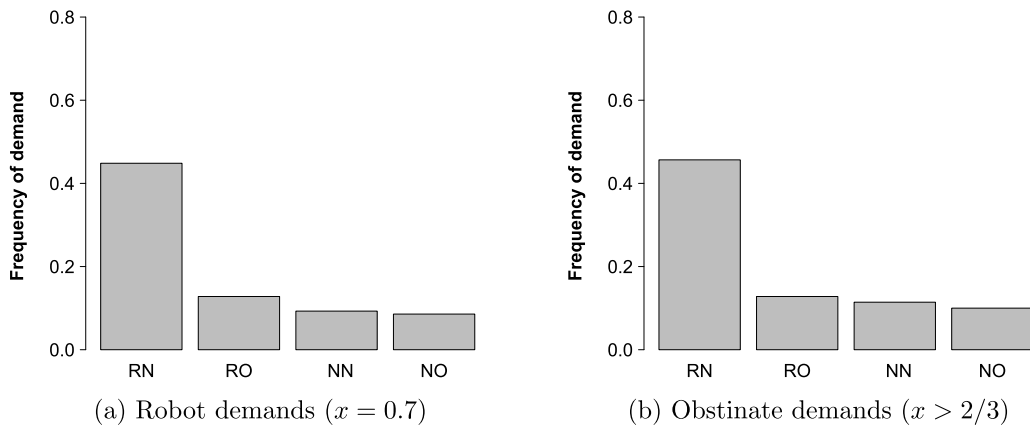
²⁴ We hold the number of subjects in a matching block constant across treatments with and without robots in order to keep the treatments as similar as possible. That is, by doing so, the only difference between treatments with and without robots is the presence of robots, *not* the presence of robots *and* the number of subjects.

²⁵ A session—defined as a collection of subjects participating in the lab on the same day and at the same time—consists of between one and three blocks.

²⁶ <https://www.socialscicenter.org/trials/2674>.

²⁷ Based on the code provided by Bellemare et al. (2016).

²⁸ In online appendix E, we present results from behavior over games. For the sake of robustness, we present results which also include data from early games in online appendix C. Including early games does not alter our results qualitatively.



Note: The first three games and demands from robots are excluded.

Fig. 1. Obstinate demands in the first period.

non-parametric (Wilcoxon) rank-sum tests using matching block averages as units of observation (p^n) and, second, p -values from parametric treatment regressions with robust standard errors clustered at the level of matching blocks (p^p). All tests are two-sided. Details from the tests are found in the online appendix B.

4.1. Imitation behavior

Result 1. *The share of first-period demands at 0.7 is substantially higher in RN than in RO.*

Consistent with theory, the share of subjects claiming 0.7 (the robot demand) of the pie is considerably lower in RO than in RN, as seen in Fig. 1a. When outside options are introduced, the share of first-period claims at 0.7 of the pie drops by a full 32 percentage points (from 45 to 13 percent). Furthermore, this treatment difference is highly significant ($p^n = 0.0012$; $p^p < 0.0001$).²⁹ Thus, with induced obstinate types, incentives in the form of suitably calibrated outside options are effective in deterring imitation.

Result 2. *The share of first-period demands at 0.7 is substantially higher in RN than in NN and NO.*

Comparing RN with the remaining treatments, we note that the frequency of demands at 0.7 is more than four times higher than in NN ($p^n = 0.0070$; $p^p = 0.0001$), and in NO ($p^n = 0.0082$; $p^p < 0.0001$). As is clear from panel b in Fig. 1, results do not change much if we consider all demands above the cut-off for a binding option (i.e., demands above 2/3 of the pie). We conjecture that this pattern, particularly the difference between NN and RN, can be explained by the failure of bargainers to coordinate their beliefs on one distribution of obstinacy. This conjecture is explored further in treatments NNc, RNc, and RNr; see section 5.

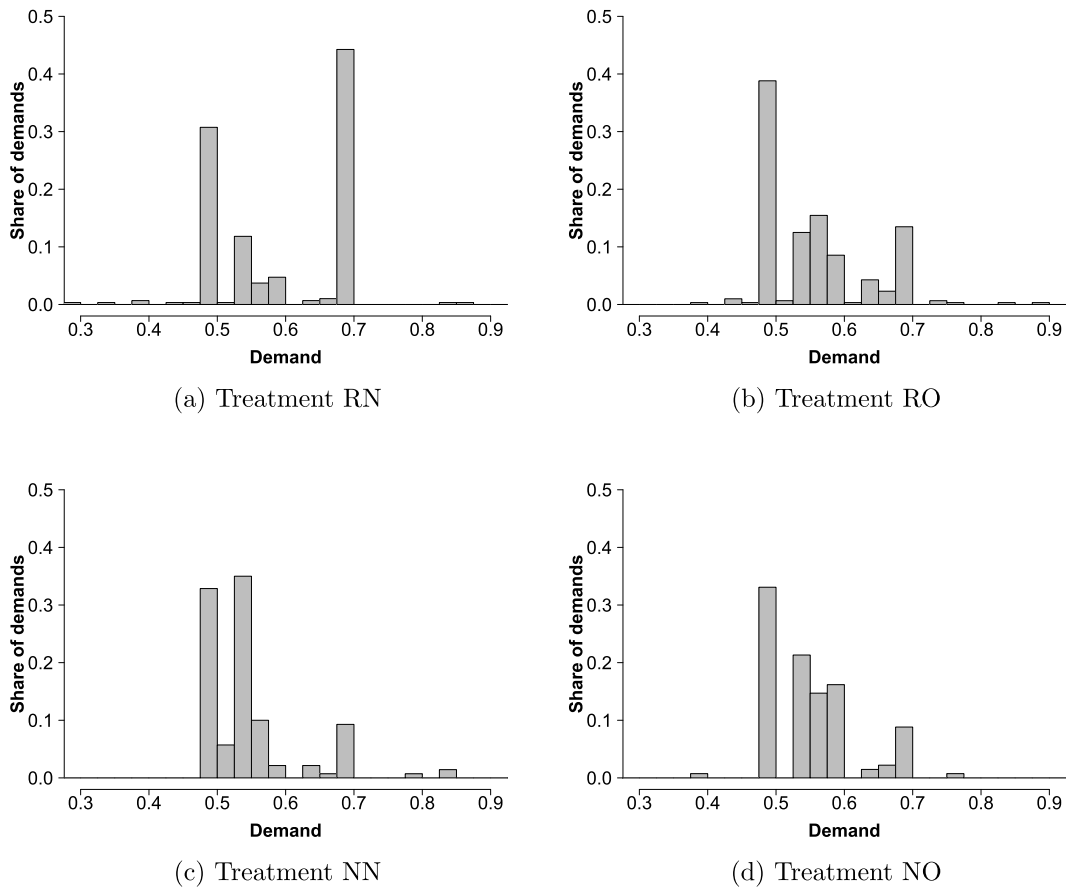
While there is a difference in obstinate demands above 2/3 between RN and RO, there is no such difference between NN and NO. The share of demands above 2/3 is 0.10 in NO and 0.11 in NN. The absence of a clear treatment effect between NN and NO could be indicative of the effect of introducing an outside option of less than 0.3 in a treatment with robots: To the extent that subjects fail to coordinate on beliefs of obstinacy above 0.7 (the robot demand), an outside option of less than 0.3 would likely have a limited effect. In section 4.2, we nuance this discussion by considering the potential role and impact of inequality aversion in the context of outside options.

Result 3. *The estimated share of obstinate human bargainers is in the range 10 to 13 percent.*

Fig. 2 displays the distribution of demands for each treatment. Focus on first-period demands in RO. Almost all demands above the level at which the outside option becomes binding result in the responder opting out (see Fig. 3b). Given that rational proposers correctly foresee this, such claims must come from irrational subjects.

The share of first-period claims above 2/3 is in roughly the same ballpark in RO, NN, and NO (13, 11, and 10, respectively). Furthermore, the observed treatment differences are not significantly different (RO vs. NN: $p^n = 0.4908$; $p^p = 0.7722$, RO vs. NO: $p^n = 0.4887$; $p^p = 0.4022$, and NN vs. NO $p^n = 0.6723$; $p^p = 0.8280$). Given that the fraction of

²⁹ The very low p -values indicate that our power calculations worked, and that the probability of a false positive is less than 20 percent for this treatment difference. Benjamin et al. (2018) provide the argument for this statement.



Note: The first three games and demands from robots are excluded. Bin size = 0.025.

Fig. 2. Histogram of first-period demands.

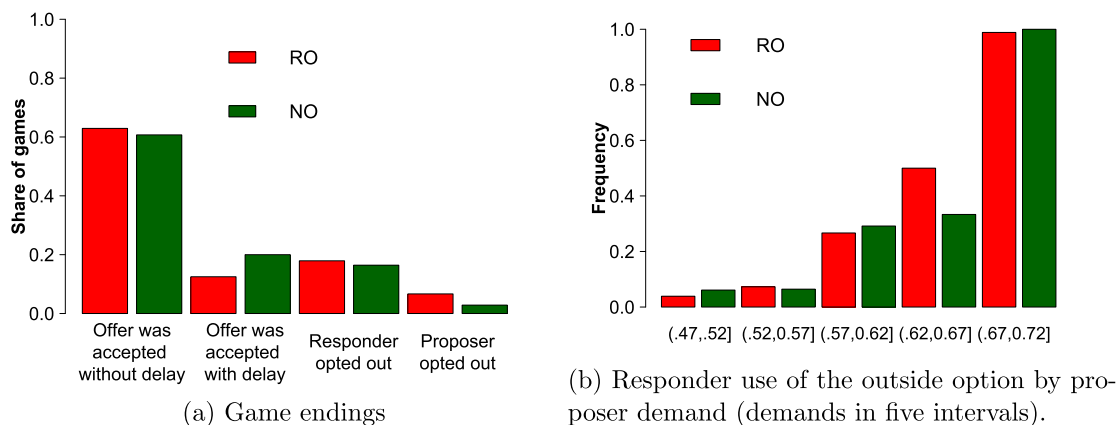
claims above $2/3$ is stable over treatments *RO*, *NN*, and *NO*, a reasonable estimate of the share of homegrown obstinates in our experiment lies in the range 10 to 13 percent. In *RN*, the shares of first-period claims above $2/3$ and at 0.7 are practically identical. It follows that the difference between demands above $2/3$ in *RN* and the estimate of homegrown obstinates provides an estimate of the share of humans imitating in *RN*, which is between 33 and 36 percent. In online appendix E, we show that there are limited learning effects: Irrational demands (above $2/3$) remain high also in later games despite being routinely rejected.³⁰ This observation strengthens an interpretation of observed aggressive demanders (subjects claiming more than $2/3$) as being truly obstinate in the sense of the model.^{31,32}

Again, see the distributions of demands in Fig. 2. While extreme demands dominate in *RN*, the demands in *RO*, *NN* and *NO* are substantially more moderate, with a large share of demands around an equal split. Our interpretation is that equal division, which also happens to be close to the equilibrium with common-knowledge rationality by our design, is a salient norm in the absence of induced obstinate types. As such, our results illustrate that the fragility of the 50-50 norm is conditional. A sizable fraction of subjects do not hesitate in deviating from the norm, which is both effective and equitable

³⁰ In online appendix E (Figure E2), we show that the share of demands above $2/3$ stabilizes after three games. We also show (Table B7) that the share of irrational demands is high and significantly larger than zero in *RO* and *NO*.

³¹ We interpret demands in excess of $2/3$ in treatments *RO* and *NO* as coming from homegrown obstinate subjects. They could, however, just as well be coming from players who are irrational in the sense of a level-0 player in a k-level model. To address the issue of the motivation behind these demands, we have looked at demands made in period 3 in games that start out with a demand in excess of $2/3$ in treatment *NN*. We look at *NN* because the share of demands in excess of $2/3$ is at the same level as in *RO* and *NO*, which indicates that there is no or very little mimicking, and because a larger number of games go beyond the first period than in *NO* and *RO*). It turns out that half of the subjects that start out with a demand in excess of $2/3$ also make a demand in excess of $2/3$ if the game reaches period 3 (where they can again make a demand). As such, our interpretation seems to be at least partly justified.

³² Although our focus is on first-period demands, we also report counteroffers made in period 2 (for games where period 2 is reached) in online appendix F. The distributions of counteroffers are very similar to those in Fig. 2 with respect to demands at 0.7 .



Note: Matches that include robots and the first three games are excluded.

Fig. 3. Use of the outside option. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

if used, provided there is an induced obstinate bargainer to imitate and no deterrence in the form of outside options (i.e., in *RN*).

Note also from Fig. 1 that demands at 0.7 make up a large share of demands only in the *RN* treatment, despite the button for a 0.7 demand being present in all treatments. This indicates that the focal point created by the mere presence of the 0.7 button is not strong enough to override the incentives created by outside options (in *RO* and *NO*), or to substitute for the lack of induced obstinate bargainers (in *NN* or *NO*).

Result 4. The frequency of demands in the range (0.5, 2/3] is significantly higher in *RO*, *NN*, and *NO* than in *RN*.

In *RO*, *NO*, and *NN*, there is a sizable mass of demands in the range (0.5, 2/3] where imitation of obstinate types could make sense for suitably specified beliefs about obstinacy: 42, 56, and 54 percent, respectively, compared to a share of only 21 percent in *RN*.³³ The differences in the shares of such demands between *RN* and the three remaining treatments are all statistically significant at conventional levels (*RN* vs. *RO*: $p^n = 0.0031$; $p^p = 0.0052$, *RN* vs. *NN*: $p^n = 0.0095$; $p^p = 0.0001$, and *RN* vs. *NO* $p^n = 0.0569$; $p^p = 0.0173$). From Fig. 2, we appreciate that in *RO*, *NN*, and *NO* the frequency of demands in the range (0.5, 2/3] is more concentrated away from 0.5 than in *RN*. Interpreting this under the assumptions of the model, subjects in *RO*, *NO*, and *NN* are either homegrown obstinates in the range (0.5, 2/3], or rational subjects imitating various forms of obstinacy that each could form an equilibrium of the model for suitably specified beliefs. For subjects in *RN*, we surmise that a common belief about obstinacy has been effectively induced and that subjects act on this belief in a way consistent with the equilibrium of the model.

4.2. Outside options and delay

Result 5. Most subjects opt out when offered less than their outside option ($1 - x < 1/3$).

See Fig. 3. Even when matches that include robots are excluded, a high share of games end by bargainers opting out. Furthermore, the lion's share of opting out is due to responders. The fraction of proposers opting out in the first period is small (below 4 percent of matches in *NO* and below 6 percent of matches in *RO*). The condition for proposers immediately taking the outside option ($v^{out} > (1 - \epsilon)v^* + \epsilon\delta v^{out}$) does not bind for our estimate of only homegrown obstinates (*NO*) or for the combined share of homegrown and robots (*RO*). This holds for pure self-regard.³⁴ It also holds for aversion to unfavorable inequality.^{35,36} Thus, the absence of proposers opting out immediately can be rationalized within the model.

³³ This share is either stable or increasing over games; see Figure E3 in online appendix E.

³⁴ Using the parameters of our experiment, we find that $\epsilon > .856$ is required for purely self-serving proposers to opt out immediately. We have .25 obstinate robots in our robot treatment, and an estimate of .10-.13 homegrown obstinates. Thus, the maximum estimated ϵ in our data is .38, well below the requirement.

³⁵ Aversion to unfavorable inequality can be viewed as a special case of Fehr and Schmidt (1999), with $\alpha \geq 0$ (the utility weight on unfavorable inequality) and $\beta = 0$ (the utility weight on favorable inequality). Alternatively, it can be viewed as a special case (a linear version of the original ERC, negative reciprocity specification) of Bolton and Ockenfels (2000). Evidence on the reasonableness of such a specification is provided by De Bruyn and Bolton (2008).

³⁶ For inequality-averse proposers, replace v^* with $u^* = \frac{1+\alpha}{1+2\alpha+\beta}$ (the equilibrium share of a proposer with aversion to unfavorable inequality; see Kohler and Schlag (2019), equation (5)). The typical distribution of inequality has $\alpha_i \leq 1$ for 90 percent of subjects; see Fehr and Schmidt (1999), Table III. Bellemare et al. (2008) replicate this finding. Thus, for 90 percent of subjects, an $\epsilon \in [0.847, 0.856]$ would provoke immediate opting out. Again, this is well above the estimated fraction of obstinates in our treatments (and very similar to the requirement for purely self-serving subjects).

Under the assumptions of the model, immediate acceptance of the equilibrium offer in the standard model ($v^* = 0.53$) should be the outcome in matches with two rational bargainers that hold a shared belief that any obstinate type will insist on 0.7 of the pie. Immediate agreement is also the most commonly observed outcome in our data, accounting for 63 percent of outcomes in *RO* and 61 percent of outcomes in *NO*. Furthermore, average demands accepted without delay are close to the equilibrium outcome of the model: 0.53 in *RO* and 0.53 in *NO*. So, in broad strokes, the data seem to provide support in favor of the model with respect to agreements.

A challenge is the sizable share of responders that opt out immediately (18 percent of matches in *RO* and 16 percent in *NO*). In online appendix D, we run separate logistic regressions for *RO* and *NO*, respectively. The dependent variable is a binary variable coded 1 for opting out and otherwise zero, which is regressed on proposers' demands. These regressions show that greedy demands account for the lion's share of opting out. In particular, for demands in excess of the outside option, the responder (almost) certainly opts out. Also, the probability curve of the responders is less steep in *NO* than in *RO*. We ascribe this to the induced obstinate types in *RO*, which helps subjects coordinate on a shared belief about obstinacy.

Result 6. *A substantial share of subjects opt out when offered more than their outside option ($1 - x > 1/3$).*

In Fig. 3b, we observe that a sizable share of responders opt out in cases where doing so leads to strictly lower material payoffs.³⁷ Thus, a share of subjects are willing to forgo material payoffs facing unfavorable demands in an unequal split. This result is consistent with results from experiments on the ultimatum game.³⁸ However, as demands move toward an equal split, the proportion of subjects opting out decreases monotonically. Such behavior is consistent with a fraction of our subject sample having a sufficiently strong aversion to unfavorable inequality. By exercising the outside option in the face of a demand larger than v^* , the subject immediately realizes an equal split in which each bargainer obtains a third of the pie. By accepting a demand greater than half the pie, subjects with an aversion to unfavorable inequality incur a utility loss.

To clarify, consider a utility function for the responder equal to $U_R(1 - x, x) = 1 - x - \alpha \max\{2x - 1, 0\}$, in which x represents the proposer's demand. Furthermore, assume that $\alpha \geq 0$ is common knowledge and define $u^* = \frac{1+\alpha}{1+2\alpha+\delta}$ as the equilibrium demand with common-knowledge rationality and (possibly) aversion to unfavorable inequality (Kohler and Schlag (2019)). With aversion to unfavorable inequality ($\alpha > 0$), the responder prefers to opt out rather than to accept the demand x if $v^{out} > 1 - x - \alpha \max\{2x - 1, 0\}$. Thus, if the responder believes that $1 - x$ is the best offer he will ever receive and that the proposer will never accept anything less than x , his best response must be to opt out immediately. If $x = \theta_m$ and $\varepsilon_m > 0$, this is exactly what he should believe. The reasoning is the same as in Compte and Jehiel (2002): Since the demand x will never be accepted, there is nothing to gain from imitating an obstinate type characterized by such a demand. Consequently, the demand x can come from only an obstinate type. As a result, while the outside option in itself deters imitation of types in the range $(1 - v^{out}, 1]$, outside options, together with an aversion to unfavorable inequality, deter imitation of types in the range $(\frac{1-v^{out}+\alpha}{1+2\alpha}, 1]$. Since $\frac{1-v^{out}+\alpha}{1+2\alpha} < 1 - v^{out}$, the outside option will deter imitation of a wider range of types when players dislike unfavorable inequality.³⁹ Note that even with aversion to unfavorable inequality, we should expect imitation of types in the range $(u^*, \frac{1-v^{out}+\alpha}{1+2\alpha}]$.

As noted, the average accepted first-period demand is 0.53 in both *RO* and *NO*. With the parameters of our experiment, u^* goes from 0.53 for $\alpha = 0$ to 0.50 as $\alpha \rightarrow \infty$. Because 90 percent of subjects are estimated to have $\alpha \leq 1$ (Fehr and Schmidt (1999) and Bellemare et al. (2008)), we should expect accepted demands around 0.53. In contrast to this, the average demand provoking immediate execution of the outside option is 0.65 in *RO* and 0.61 in *NO*. We surmise that the observed pattern of immediate agreements and immediate execution of the outside option is qualitatively in line with a model that allows for a typical distribution of unfavorable inequality aversion in our subject sample. Thus, assuming typical fairness preferences goes a long way in rationalizing the excessive use of outside options by responders compared to the prediction of Compte and Jehiel (2002), which is based on purely self-regarding behavior.⁴⁰

In addition to implementing an equal split, exercising the outside option in the face of inequitable claims inflicts harm on the opponent. This is valued in and of itself by strong reciprocators (e.g., Fehr and Gächter (2000) and Charness and Rabin (2002)). If a fraction of subjects in our sample are also strong reciprocators, outside options are inflated in utility terms, contributing to the excessive use of them.⁴¹

Result 7. *There is significantly less delay in treatments *RO* and *NO* than in treatments *RN* and *NN*.*

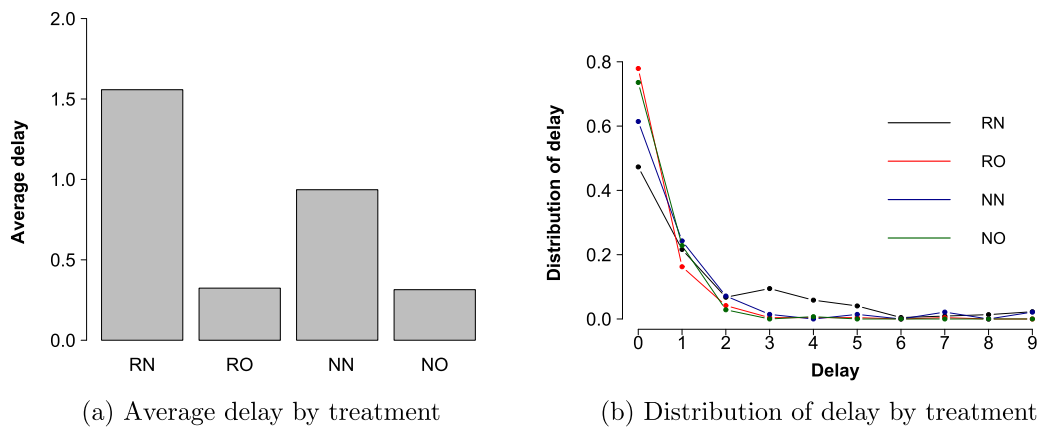
³⁷ This type of behavior is inconsistent with any belief a subject might have about the opponent following such a demand. That is, even if a subject believes that this demand comes from some obstinate type, accepting the offer dominates opting out in terms of material payoffs. Thus, we must look beyond explanations such as non-Bayesian updating to account for this behavior.

³⁸ See, e.g., Fehr and Schmidt (1999) and Bellemare et al. (2008).

³⁹ Note that aversion to unfavorable inequality will not lead to the use of the outside option in the absence of obstinate types. That is, even when α is high, there always exists a split of the pie that is strictly preferred to opting out by both players when $\varepsilon = 0$.

⁴⁰ Embrey et al. (2015) observe a fraction of subjects with initial claims of 1/3. They interpret this as a fraction of subjects having preferences for "conflict avoidance." Opting out does not easily lend itself to such an interpretation in our experiment, since very few proposers in our data use their outside option in the first period.

⁴¹ Bruhin et al. (2018) provide detailed evidence on the distribution of strong reciprocators in the pool of experimental subjects.



Note: Matches that include robots and the first three games are excluded.

Fig. 4. Bargaining delay. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

Due to the behavior discussed above, there is a substantial treatment effect on delay of having access to an outside option in both the treatments with and without robots. See Fig. 4a and focus first on the robot treatments where it is plausible to assume that bargainers’ beliefs have converged on obstinates and their imitators demanding 0.7 of the pie. Games are on average delayed by 1.6 periods in RN compared to only 0.3 periods in RO ($p^n = 0.0003$; $p^p < 0.0001$). This difference in delay is largely driven by parties immediately accepting offers rather than opting out in RO. If games that terminate by subjects opting out are excluded, the average delay in RO drops to only 0.2.⁴² Comparing RN and NN, we note that there is a statistically uncertain increase in the average delay of 0.6 periods. While the observed increase is significantly different from zero in a parametric test, it does not reach the 5% threshold in a non-parametric test ($p^n = 0.1262$; $p^p = 0.0489$).

We have ascribed the stronger effect in robot treatments to a shared belief about obstinacy that pushes imitation and delay more forcefully in RN than in NN. The sign of this treatment effect is consistent with the one reported in Embrey et al. (2015). Moreover, it is well known that diverging beliefs on which principles it is reasonable to play by will create an impasse and a breakdown in unstructured bargaining (e.g., Roth and Murnighan (1982), and Babcock and Loewenstein (1997)). Our results suggest that a similar finding holds for reputation-building in alternating-offer bargaining.

There is a fraction of bargaining situations that end in delayed agreements when outside options are available (16 percent in RO and 23 percent in NO, respectively). By insisting on an equal division, the responder enters an uncertain world. Prolonged delay may well result from being matched with an induced (in RO) or homegrown obstinate (in RO and NO).

With a discount factor of 0.9, as in the experiment, obtaining a third of the pie immediately is worth more in purely monetary terms than acceptance of half of the remaining pie later than period four. Thus, from a strict monetary perspective, if subjects expect a degree of persistence of a demand larger than v^* from their opponent, payoffs are maximized by opting out early. This incentive is further strengthened if subjects have fairness preferences or are reciprocating.

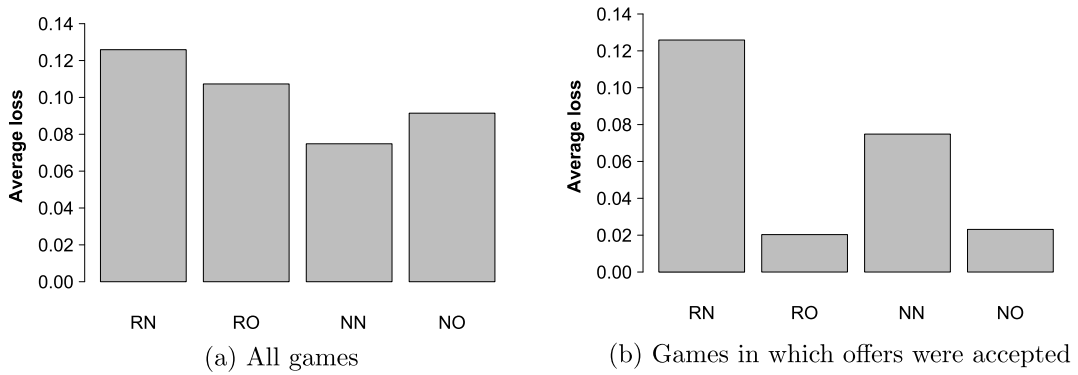
Result 8. *There is no significant effect of outside options on efficiency.*

The availability of outside options drives down imitation and delay, which in turn would suggest that it should generate efficiency. While the partial effect of these two factors does reduce the amount of surplus that is lost due to discounting, this loss is offset by the excessive use of outside options. As such, there is no effect on overall efficiency; see Fig. 5a. The difference in lost surplus between RN and RO is 2 percentage points ($p^n = 0.4517$; $p^p = 0.3770$). If we exclude games in which bargainers opted out and consider only games in which an offer was eventually accepted, the average share of the pie that is lost in RO drops from 11 to 2 percent. Thus, the use of the outside option accounts for more than 80 percent of the lost surplus in this treatment and the outside option does reduce the loss induced by delay by a substantial amount. Considering the other treatments, we note that induced obstinacy has an effect on efficiency. Comparing NN and RN we observe a significant difference in lost surplus of 6 percentage points ($p^n = 0.0495$; $p^p = 0.0190$).

5. Additional treatments

Comparing demands across NN and NO, we have noted that there is no significant difference in the share of obstinate demands (demands above 2/3). Taking the predictions from Compte and Jehiel (2002), this similarity of demands is puzzling

⁴² Also worth noting is that the delay in NN is substantially higher than both the delay in RO, with a treatment difference of 0.6 periods, and the delay in NO, with a treatment difference of 0.6 periods ($p^n = 0.0176$; $p^p = 0.0116$, and $p^n = 0.0119$; $p^p = 0.0114$).



Note: Matches that include robots and the first three games are excluded.

Fig. 5. Loss of surplus.

because we would expect imitation of obstinacy to take place in *NN*: There is a considerable share of first-period demands above $2/3$ in *NO*, and these irrational demands provide us with an estimate of homegrown obstinates in our sample. Given this estimate, the difference between *NN* and *NO* provides us with a measure of imitation by rational players of such obstinate types in *NN*. Even though a share of obstinate demands indeed is present in *NN*—and there is in principle a large set of commonly held beliefs about obstinacy that would be consistent with imitation behavior as part of equilibrium play—we observe no significant imitation according to our measure.

In our view, the challenge for bargainers is to coordinate their beliefs on one distribution of obstinacy. We conjecture that the absence of induced obstinate types makes it unlikely that such a common belief will emerge. While the robots in *RN* and *RO* induce a convention, no such inducement is present in *NN* and *NO*. Thus, it seems plausible that outside options are a necessary but not sufficient condition for the core model prediction to manifest itself in the data. We surmise that deterrence of imitation in bargaining requires both a commonly held belief about obstinacy and suitably calibrated outside options.

The last three treatments are designed to explore the role of induced obstinate types in shaping beliefs. These treatments are without outside options. First we constrain the strategy set to $\{0.5, 0.7\}$ to obtain experimental control over the way a subject may be obstinate or imitate obstinates. The constrained treatment *RNc* is with robots, and the difference in the share of obstinate demands between *RNc* and *RN* indicates the effect of constraining the strategy set. The constrained treatment *NNc* is without robots, and the difference between *NNc* and *NN* indicates whether constraining the choice set can substitute for robots in shaping beliefs about obstinacy. Severely constraining the choice set like this ought to facilitate coordination on a shared belief about what constitutes obstinacy. After all, it is now one, and only one, way of being obstinate. Provided coordination on a shared belief about obstinacy is necessary for imitation to occur, the constrained choice set should at least not lead to a fall in imitation behavior compared to the unconstrained choice set. Note that we do not measure beliefs directly in any of our treatments. As such, our results should be viewed as indicative of the role of shared beliefs.

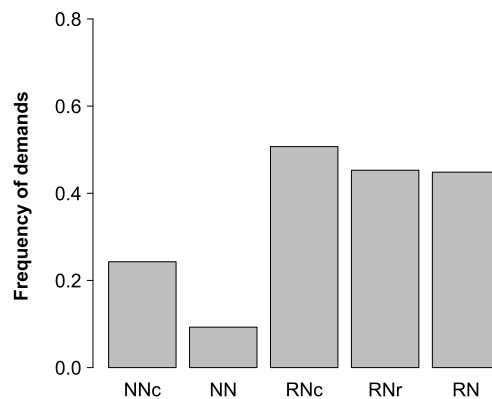
The last treatment—*RNr*—does not constrain the strategy set, but reveals to subjects whether the match was a robot or a human subject after the termination of each bargaining game.⁴³ Note that this information is revealed to only the subjects *within* the match. The difference in the share of obstinate demands between *RNr* and *RN* indicates the effect of imposing a social-image cost of imitating an obstinate player.⁴⁴ If the role of robots is mainly to offer subjects a chance device to hide behind, and thus a way of escaping social-image costs, we should see a drop in imitation behavior when going from *RN* to *RNr*. If there is no significant drop in imitation behavior, it indicates that robots mainly serve to coordinate beliefs, rather than to reduce social-image costs. Apart from the noted differences, the protocol and instructions in the last three treatments are identical to the ones used in the first four treatments of the experiment.

Consider Fig. 6. In *RNc* and *NNc*, there is only one way in which to be obstinate, corresponding to the obstinate types at $\theta = 0.7$ of the model. A priori, one would at least expect an increase in the share of obstinate demands in the constrained treatments for purely mechanical reasons (i.e., any demand deviating from an equal split must by design correspond to the obstinate demand at 0.7).

Comparing *RN* and *RNc*, we do indeed observe an increase of 6 percentage points in obstinate demands at 0.7 when moving to the constrained treatment. This difference, however, is (far from) significant at conventional levels ($p^n = 0.7515$; $p^p = 0.4100$). Why do we fail to observe an increase in such obstinate demands in *RNc*? Our favored interpretation is that

⁴³ Also, this aspect of the implementation was made public knowledge by handing out and reading aloud the instructions.

⁴⁴ The term “social-image costs” is used in Abeler et al. (2019) to capture the sentiment that “...the decision maker feels bad to the extent that the audience believes he cheats” (Dufwenberg and Dufwenberg, 2018).



Note: Robot demands and the first three games are excluded.

Fig. 6. Obstinate demands at 0.7 in the first period.

robots induce shared beliefs about obstinacy effectively, and that constraining the set of available demands does little to further strengthen these beliefs.

Comparing *NN* and *NNc*, the picture changes somewhat. In the absence of robots, we observe an increase of 15 percentage points in obstinate demands at 0.7 when available demands are constrained. However, this increase is significant only in the parametric test ($p^n = 0.1094$; $p^p = 0.0090$). Since the effect of constraining the set of available demands is absent in the presence of robots, we are reluctant to ascribe the observed effect in the absence of robots to a purely mechanical effect. A more likely explanation for the observed difference between *NN* and *NNc* is that constraining the set does not succeed in forcefully shaping beliefs about obstinacy. In our bargaining environment, inducing obstinate types with robots seems to be a more effective way of establishing shared beliefs than constraining the set of available demands is.⁴⁵ Another possible explanation as to why we observe limited effects of constraining the choice set is that 0.7 becomes the only aggressive demand. That is, if subjects believe that others reciprocate based on the ranking of demands, they might be reluctant to demand 0.7 out of fear of negative reciprocity. Had we, say, included an option of demanding 0.9, demands at 0.7 could be believed to be ranked more favorably and might therefore be chosen more often.⁴⁶

A reason why robots might be more effective in establishing shared beliefs is hinted at in the experimental literature. A replicable finding from ultimatum bargaining is that the frequency of greedy demands increases when proposers can hide behind chance devices that provide them with private information about the pie size (Kagel et al. (1996), Guth (1995), Mitzkewitz and Nagel (1993), Rapoport and Sundali (1996)).⁴⁷ Furthermore, a large and growing experimental literature shows that a sizable fraction of subjects in non-strategic experiments are willing to perform profitable lies provided such behavior remains undetected (Abeler et al. (2019), Gneezy et al. (2018)). In the context of our experiment, if subjects are aware of such tendencies in others, they should place more probability weight on the opponent's willingness to imitate robots, on the opponent's beliefs about the subject's willingness to imitate, and so on. This would create a momentum in the direction of shared beliefs about imitation of obstinacy in an environment where robots can be profitably imitated. Merely constraining the set of available demands (as in *NNc*) does not allow subjects to mask their actions convincingly, and will not create the same momentum in the direction of shared beliefs.

Our last treatment—*RNr*—reveals to the opponent (the audience) whether it faced a robot or a human subject. This revelation is publicly known to happen after the termination of a bargaining game. While type (human or robot) is revealed, subject-to-subject anonymity is preserved.⁴⁸ If the main channel through which robots work is the reduction in social-image costs of deceiving, revealing attempts at deception to the opponent should be expected to drive down imitation. However, the treatment difference ($RN - RNr$) is less than one percentage point, and this difference is not significant at conventional levels ($p^n = 0.792$; $p^p = 0.8768$). We interpret this as evidence that robots work mainly through the shaping of shared beliefs that can be acted on consistently, rather than by lowering the social-image costs of deception.

⁴⁵ Like us, Embrey et al. (2015) find substantially more imitation of high demands (above an equal split) in treatments with robots than in treatments with no robots. In contrast to us, however, they observe an increase of high demands with constrained treatments in the presence of robots. There are many important differences between our bargaining environment and theirs. For example, in their study, subjects play a Nash demand game and enter a war of attrition if demands are incompatible, while we use an infinite alternating-offer protocol. Moreover, while they pay subjects by the lottery procedure, we use monetary payoffs.

⁴⁶ There is experimental evidence to support the existence of such choice-set effects: see Brandts and Solà (2001), Falk et al. (2003), and Owens and Kagel (2010).

⁴⁷ Similar findings hold for dictator giving as well (Dana et al. (2007), Larson and Capra (2009)).

⁴⁸ Imposing administrator-subject anonymity in individual decision-making experiments seems to increase lying (Abeler et al., 2019). Little is known about the effects on imitation in strategic decision-making. However, and somewhat relatedly, Larson and Capra (2009) find no effect of varying such anonymity on the propensity to costlessly reveal information about payoff consequences in the framework of Dana et al. (2007).

6. Conclusion

The results from the experiment are broadly consistent with the theory. The introduction of outside options leads to a drop in imitation behavior and delay. However, we find the power of outside options to be contingent on induced obstinacy. That is, we find no treatment effect of outside options on imitation in the absence of robot players. These results indicate that for imitation to take place in bargaining, there need to be shared beliefs about obstinacy as well as an absence of outside options.

Moreover, we observe a sizable share of (irrationally) high demands that are met by responders opting out. This also happens when responders are faced with demands that would leave them with a higher material payoff than taking their outside option would. We argue that this is consistent with fairness preferences being present in our subject sample. As a result, outside options do not improve on efficiency in our experiment even though they powerfully prevent delay.

Declaration of competing interest

The authors have received financial support from the Research Council of Norway, grant #250506, for the research culminating in the paper “The Power of Outside Options in the Presence of Obstinate Types”.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.geb.2022.10.011>.

References

- Abeler, J., Nosenzo, D., Raymond, C., 2019. Preferences for truth-telling. *Econometrica* 87, 1115–1153.
- Abreu, D., Gul, F., 2000. Bargaining and reputation. *Econometrica* 68, 85–117.
- Abreu, D., Pearce, D., 2007. Bargaining, reputation, and equilibrium selection in repeated games with contracts. *Econometrica* 75, 653–710.
- Abreu, D., Sethi, R., 2003. Evolutionary stability in a reputational model of bargaining. *Games Econ. Behav.* 44, 195–216.
- Anderhub, V., Engelmann, D., Güth, W., 2002. An experimental study of the repeated trust game with incomplete information. *J. Econ. Behav. Organ.* 48, 197–216.
- Atakan, A.E., Ekmekci, M., 2013. Bargaining and reputation in search markets. *Rev. Econ. Stud.* 81, 1–29.
- Babcock, L., Loewenstein, G., 1997. Explaining bargaining impasse: the role of self-serving biases. *J. Econ. Perspect.* 11, 109–126.
- Bellemare, C., Bissonnette, L., Kröger, S., 2016. Simulating power of economic experiments: the powerBBK package. *J. Econ. Sci. Assoc.* 2, 157–168.
- Bellemare, C., Kröger, S., Van Soest, A., 2008. Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities. *Econometrica* 76, 815–839.
- Benjamin, D.J., Berger, J.O., Johannesson, M., Nosek, B.A., Wagenmakers, E.-J., Berk, R., Bollen, K.A., Brembs, B., Brown, L., Camerer, C., et al., 2018. Redefine statistical significance. *Nat. Hum. Behav.* 2, 6.
- Berg, J.E., Rietz, T.A., Dickhaut, J., 2008. On the performance of the lottery procedure for controlling risk preferences, 1 ed., vol. 1. Elsevier, pp. 1087–1097. Part 8, chap. 115.
- Binmore, K., Shaked, A., Sutton, J., 1989. An outside option experiment. *Q. J. Econ.* 104, 753–770.
- Bochet, O., Siegenthaler, S., 2018. Better later than never? An experiment on bargaining under adverse selection. *Int. Econ. Rev.* 59, 947–971.
- Bolton, G.E., Ockenfels, A., 2000. ERC: a theory of equity, reciprocity, and competition. *Am. Econ. Rev.* 90, 166–193.
- Brandts, J., Solà, C., 2001. Reference points and negative reciprocity in simple sequential games. *Games Econ. Behav.* 36, 138–157.
- Brekke, K.A., Ciccone, A., Heggedal, T.-R., Helland, L., 2022. Reference points in sequential bargaining: Theory and experiment. Working Paper.
- Bruhlin, A., Fehr, E., Schunk, D., 2018. The many faces of human sociality: uncovering the distribution and stability of social preferences. *J. Eur. Econ. Assoc.* 17, 1025–1069.
- Charness, G., Rabin, M., 2002. Understanding social preferences with simple tests. *Q. J. Econ.* 117, 817–869.
- Chatterjee, K., Samuelson, L., 1987. Bargaining with two-sided incomplete information: an infinite horizon model with alternating offers. *Rev. Econ. Stud.* 54, 175–192.
- Chatterjee, K., Samuelson, L., 1988. Bargaining under two-sided incomplete information: the unrestricted offers case. *Oper. Res.* 36, 605–618.
- Coase, R.H., 1972. Durability and monopoly. *J. Law Econ.* 15, 143–149.
- Compte, O., Jehiel, P., 2002. On the role of outside options in bargaining with obstinate parties. *Econometrica* 70, 1477–1517.
- Cramton, P.C., 1984. Bargaining with incomplete information: an infinite-horizon model with two-sided uncertainty. *Rev. Econ. Stud.* 51, 579–593.
- Crawford, V.P., 1982. A theory of disagreement in bargaining. *Econometrica*, 607–637.
- Dana, J., Weber, R.A., Kuang, J.X., 2007. Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Econ. Theory* 33, 67–80.
- De Bruyn, A., Bolton, G.E., 2008. Estimating the influence of fairness on bargaining behavior. *Manag. Sci.* 54, 1774–1791.
- Dufwenberg, M., Dufwenberg, M.A., 2018. Lies in disguise – a theoretical analysis of cheating. *J. Econ. Theory* 175, 248–264.
- Ekmekci, M., Zhang, H., 2020. Reputational Bargaining in the Shadow of the Law. Available at SSRN 3262107.
- Ellingsen, T., Miettinen, T., 2008. Commitment and conflict in bilateral bargaining. *Am. Econ. Rev.* 98, 1629–1635.
- Ellingsen, T., Miettinen, T., 2014. Tough negotiations: bilateral bargaining with durable commitments. *Games Econ. Behav.* 87, 353–366.
- Embrey, M., Fréchette, G.R., Lehrer, S.F., 2015. Bargaining and reputation: an experiment on bargaining in the presence of behavioural types. *Rev. Econ. Stud.* 82, 608–631.
- Falk, A., Ernst, F., Urs, F., 2003. On the nature of fair behavior. *Econ. Inq.* 41, 20–26.
- Fanning, J., 2016. Reputational bargaining and deadlines. *Econometrica* 84, 1131–1179.
- Fanning, J., Kloosterman, A., 2022. An experimental test of the Coase conjecture: fairness in dynamic bargaining. *Rand J. Econ.* 53, 138–165.
- Fehr, E., Gächter, S., 2000. Fairness and retaliation: the economics of reciprocity. *J. Econ. Perspect.* 14, 159–181.
- Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition, and cooperation. *Q. J. Econ.* 114, 817–868.
- Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. *Exp. Econ.* 10, 171–178.
- Friedenberg, A., 2019. Bargaining under strategic uncertainty: the role of second-order optimism. *Econometrica* 87, 1835–1865.

- Fudenberg, D., Tirole, J., 1983. Sequential bargaining with incomplete information. *Rev. Econ. Stud.* 50, 221–247.
- Gneezy, U., Kajackaite, A., Sobel, J., 2018. Lying aversion and the size of the Lie. *Am. Econ. Rev.* 108, 419–453.
- Greiner, B., 2015. Subject pool recruitment procedures: organizing experiments with ORSEE. *J. Econ. Sci. Assoc.* 1, 114–125.
- Grosskopf, B., Sarin, R., 2010. Is reputation good or bad? An experiment. *Am. Econ. Rev.* 100, 2187–2204.
- Gul, F., Sonnenschein, H., 1988. On delay in bargaining with one-sided uncertainty. *Econometrica* 56, 601–611.
- Guth, W., 1995. On ultimatum bargaining experiments – a personal review. *J. Econ. Behav. Organ.* 27, 329–344.
- Heggedal, T.-R., McKay, T., 2022. Bargaining and Time Preferences: an Experimental Study. Working Paper.
- Huck, S., Lünser, G.K., Tyran, J.-R., 2016. Price competition and reputation in markets for experience goods: an experimental study. *Rand J. Econ.* 47, 99–117.
- Kagel, J.H., Kim, C., Moser, D., 1996. Fairness in ultimatum games with asymmetric information and asymmetric payoffs. *Games Econ. Behav.* 13, 100–110.
- Kim, J., Lim, W., Schweighofer-Kodritsch, S., 2021. Patience Is Power: Bargaining and Payoff Delay. Tech. Rep.
- Kohler, S., Schlag, K.H., 2019. Inequality aversion causes equal or unequal division in alternating-offer bargaining. *Bull. Econ. Res.* 71, 47–56.
- Kreps, D.M., Wilson, R., 1982. Reputation and imperfect information. *J. Econ. Theory* 27, 253–279.
- Larson, T., Capra, C.M., 2009. Exploiting moral wiggle room: illusory preference for fairness? A comment. *Judgm. Decis. Mak.* 4, 467.
- Milgrom, P., Roberts, J., 1982. Predation, reputation, and entry deterrence. *J. Econ. Theory* 27, 280–312.
- Mitzkewitz, M., Nagel, R., 1993. Experimental results on ultimatum games with incomplete information. *Int. J. Game Theory* 22, 171–198.
- Myerson, R.B., 1979. Incentive compatibility and the bargaining problem. *Econometrica* 47, 61–73.
- Myerson, R.B., 1991. *Game Theory - Analysis of Conflict*. Harvard University Press.
- Ochs, J., Roth, A., 1989. An experimental study of sequential bargaining. *Am. Econ. Rev.* 79, 355–384.
- Owens, M.F., Kagel, J.H., 2010. Minimum wage restrictions and employee effort in incomplete labor markets: an experimental investigation. *J. Econ. Behav. Organ.* 73, 317–326.
- Özyurt, S., 2015a. Bargaining, reputation and competition. *J. Econ. Behav. Organ.* 119, 1–17.
- Özyurt, S., 2015b. Searching for a bargain: power of strategic commitment. *Am. Econ. J. Microecon.* 7, 320–353.
- Perlroth, A., 2020. Posturing in Bargaining to Influence Outsiders. Stanford GSB.
- Raiiffa, H., Richardson, J., Metcalfe, D., et al., 2002. *Negotiation Analysis: The Science and Art of Collaborative Decision Making*. Harvard University Press.
- Rapoport, A., Sundali, J.A., 1996. Ultimatums in two-person bargaining with one-sided uncertainty: offer games. *Int. J. Game Theory* 25, 475–494.
- Roth, A.E., Malouf, M.W., 1979. Game-theoretic models and the role of information in bargaining. *Psychol. Rev.* 86, 574–594.
- Roth, A.E., Murnighan, J.K., 1982. The role of information in bargaining: an experimental study. *Econometrica*, 1123–1142.
- Rubinstein, A., 1985. A bargaining model with incomplete information about time preferences. *Econometrica* 53, 1151–1172.
- Sagan, S.D., Suri, J., 2003. The madman nuclear alert: secrecy, signaling, and safety in October 1969. *Int. Secur.* 27, 150–183.
- Schelling, T.C., 1956. An essay on bargaining. *Am. Econ. Rev.* 46, 281–306.
- Schweighofer-Kodritsch, S., 2018. Time preferences and bargaining. *Econometrica* 86, 173–217.
- Sobel, J., Takahashi, I., 1983. A multistage model of bargaining. *Rev. Econ. Stud.* 50, 411–426.