Gender Differences in Bargaining with Asymmetric Information*

Samreen Malik† Benedikt Mihm‡
Maximilian Mihm§ Florian Timme¶

September 21, 2020

Abstract

We conduct an experiment on gender differences in bargaining environments with two-sided asymmetric information. Based on the bargaining model in Abreu and Gul (2000), we introduce asymmetric information about subjects' commitments by inducing irrational types, who never back down from a fixed bargaining position. We find that bargaining behavior in this setting depends on whether gender is revealed or not. When gender is unknown, men are more likely to exploit information asymmetries by mimicking irrational types than women, but this gender gap is eliminated when gender is known. Male-male pairs also experience significantly longer bargaining delays than mixed-gender or female-female pairs, but only in the treatment condition where gender is revealed and only in those pairs where one of the subjects mimics the irrational type.

Keywords: Bargaining, gender, asymmetric information, strategic posture.

JEL-Codes: J16, C78, D82.

1 Introduction

Bargaining behavior affects economic outcomes in a wide variety of wage and price negotiations. For instance, it has been widely posited that differences in the bargaining behavior of men and women could be an important factor in the persistence of a gender wage gap (see, e.g., Card et al., 2016). For example, in Germany (which provides the context for our study), the gender wage gap remains persistently high, being consistently above 21% for the last decade, with no

*We thank Matthew Embrey, Roberto Weber, and Joachim Weimann for helpful discussions, as well as an associate editor, Enrique Fatas, and two anonymous referees for very helpful comments. We are especially grateful to Guillaume Frechette for invaluable discussions, suggestions and comments.

†New York University Abu Dhabi. Email: samreen.malik@nyu.edu.
‡Magdeburg University. Email: mihm@ovgu.de.
§New York University Abu Dhabi. Email: max.mihm@nyu.edu.
¶Magdeburg University. Email: florian.timme@ovgu.de.

sign of an imminent downward trend.[1] Similarly, gender differences in bargaining behavior can affect *inter alia* the surplus that men and women acquire in family planning, divorce settlements, house or car price negotiations, and legal plea-bargains.

As a result, there is broad interest in understanding gender differences in bargaining. This is challenging, however, because bargaining outcomes (e.g., wages) can depend on many factors (e.g., productivity, sorting, or discrimination) that are not directly related to bargaining behavior. Laboratory experiments provide a setting where bargaining behavior can be isolated from such confounds (see, e.g., Azmat and Petrongolo, 2014). Accordingly, a growing experimental literature has studied gender differences in a variety of stylized bargaining problems, including dictator (Eckel and Grossman, 1998; Andreoni and Vesterlund, 2001), ultimatum (Eckel and Grossman, 2001; Solnick, 2001), and alternating-offer (Dittrich et al., 2014) games.

In this paper, we study gender differences in *strategic posturing*, a bargaining strategy that is especially relevant in dynamic bargaining problems with asymmetric information. To illustrate, consider two parties, Ann and Bob, bargaining about the division of a pie. In a symmetric, complete information environment, it seems natural for Ann and Bob to propose (and accept) an equal division of the pie, since following this 50:50 norm leads to a fair and efficient allocation. But what should Bob do if, instead, Ann initially demands a disproportionate share of the pie for herself? If Ann is irrationally obstinate about her demand—and is either unable or unwilling to deviate—then Bob should concede in order to avoid costly bargaining delays. If, on the other hand, Ann is as concerned about bargaining delays as Bob, then there is no reason why Bob should be the one to concede. The problem for Bob is that he may not be able to distinguish whether Ann is irrational or not, especially because Ann has every reason to convince Bob that she is an irrational type in order to provoke concessions. Such "bluffs" seem an essential feature of bargaining. Moreover, Abreu and Gul (2000) show that strategic posturing—whereby parties mimic irrational types—is an equilibrium strategy that can lead to significant bargaining delays and a highly unequal division of surplus.[2]

While the previous literature has identified a number of differences in the bargaining behavior of men and women, not much is known about their propensities for strategic posturing. Bargaining games studied in the prior literature—e.g., dictator, ultimatum, and alternating offer

[1]This figure is based on the unadjusted wage gap data available at Eurostat. By comparison, the EU average for the wage gap is approximately 15%. While Germany may seem a quite gender equal society—for instance, having had a female Chancellor, Angela Merkel, for over a decade—the persistently high gender wage gap indicates that there are still considerable gender differences in German labor market outcomes, which may reflect (in part) differences in bargaining behaviors of men and women.

[2]Bargaining parties may be obstinate about their demands for various reasons, not all of which are irrational: (i) parties could be committed to a pre-specified outcome because they are delegated to bargaining on someone's behalf (see, e.g., Fershtman and Kalai, 1997; Schotter et al., 2000; Fershtman and Gneezy, 2001); (ii) financial or institutional constraints could prevent parties from deviating from a pre-specified outcome (see, e.g. Genesove and Mayer, 1997); or (iii) parties could be boundedly rational, follow a simple rule-of-thumb or bargaining convention, which makes them unwilling to concede (see, e.g., Myerson, 1991).

games—share the common feature that the bargaining protocol is fundamentally asymmetric, often instilling one of the bargaining partner's with a distinct first-mover advantage. Real-world bargaining problems are often asymmetric (e.g., employer/employee or defendant/prosecutor), and prior studies provide valuable insights regarding gender differences in the ability or willingness to exploit first-mover advantage to extract more surplus from trade. However, strategic posturing does not inherently rely on any asymmetry of the bargaining protocol. Instead, strategic posturing is related to an alternative source of bargaining power: asymmetric information. To adopt a strategic posture, bargaining parties must be willing to exploit information asymmetries, "bluffing" about their commitment to a fixed bargaining position to induce concessions from their partner. Since asymmetric information is common to a wide-range of real-world bargaining problems (from wage negotiations to plea settlements), gender differences in the willingness to strategically posture can have far reaching implications: if men and women differ in their posturing behavior, this will affect the resources they acquire, and the inefficiencies they generate, from bargaining. In addition, there are reasons to believe *a priori* that there may be systematic gender differences.

On one hand, previous research provides evidence that could suggest that women may be less inclined to strategic posturing than men. First, while men often behave more in their own self-interests, women tend to be more other-regarding (Eckel and Grossman, 1998), a gender difference that may impact their willingness to demand a disproportionate share of surplus for themselves. Second, women are less likely than men to deceive for financial gain (Dreber and Johannesson, 2008), and may therefore be less inclined to feign commitment to a favorable bargaining position. Third, there is experimental evidence that women are less competitive than men (Niederle and Vesterlund, 2007), and may thus be more reluctant to pursue a competitive bargaining process.

On the other hand, a common perception that women pursue their self-interests less aggressively than men—whether founded in reality or based on gender-stereotypes—could actually provide women with favorable conditions to succeed with strategic postures.[3] If women are expected to be less aggressive in pursuing their self-interests, the conditional probability that an opponent is irrational will generally be higher when a women mimics an irrational type than when a man mimics. For instance, suppose that Charlie believes that men are generally selfish, likely to deceive for financial gain, and willing to engage in competitive interactions. In a bilateral bargaining problem, Charlie therefore anticipates that Bob will demand a disproportionate share of the pie; not because Bob is truly irrational but simply because men are generally aggressive in pursuing their self-interests. As a result, a strategic posture by Bob does

[3]There is considerable experimental evidence that people do have different prior expectations about the way that men and women behave. For instance, in a dictator game, Aguiar et al. (2009) find that subjects are more likely to choose to receive from a female dictator than a male dictator, indicating that they expect women to be more altruistic (see Section 2).

not provide a credible signal of Bob's commitment. By contrast, if Charlie believes that women are generally other-regarding, honest, and averse to competition, he may infer that Ann is truly irrational when—counter to his prior expectations—Ann demands a disproportionate share for herself. A strategic posture therefore sends a more credible signal of Ann's commitment to her bargaining position, and is more likely to provoke a concession from Charlie.

Given these countervailing forces, we conduct an experiment to investigate gender differences in strategic posturing. Our basic design is based on Embrey et al.'s (2015) implementation of the bargaining model in Abreu and Gul (2000). The underlying bilateral bargaining game has two stages. In the first stage, bargaining parties simultaneously announce what share of a pie they demand for themselves. If demands are compatible (i.e., do not exceed the total pie), then each partner receives their demanded share, any remaining pie is split equally, and the bilateral interaction ends. If the demands are incompatible, the subjects enter a second stage continuous-time concession game, where they continually decide whether to remain committed to their initial bargaining position or concede to their partner. In that case, the pie gradually shrinks until one partner concedes, and the discounted pie is then split according to the share demanded by the non-conceding party.

One challenge for studying strategic posturing in such dynamic bargaining problems is that there may be many potential irrational types in a population, making it difficult to disentangle strategic postures from other bargaining behaviors. A key innovation in Embrey et al. (2015), which we incorporate into our design, is to induce specific irrational types in the laboratory. In particular, in our experiment, some subjects are constrained to be obstinate: they must demand a disproportionate share in the first stage and can never concede in the second stage. All subjects know the likelihood of encountering one of these induced types, but do not know whether their partner is constrained or not. Inducing types in the laboratory makes the asymmetric information problem especially salient, and allows us to identify a specific strategy—mimicking the induced type—that we can directly interpret as a strategic posture.

We build on the design in Embrey et al. (2015) to investigate the role of gender and gender-revelation in a dynamic bargaining environment with two-sided asymmetric information about types. Gender seems important in such asymmetric information environments where one might expect stereotyping or statistical discrimination to have the greatest impact on bargaining behavior (Fang and Moro, 2011). While the bargaining game for our experiment confronts subjects with a more complex strategic environment than simpler bargaining games—e.g., dictator, ultimatum or alternating offer games—it offers a number of advantages for studying gender differences in posturing behavior.

First, the bargaining protocol is completely symmetric, so that willingness to exploit information asymmetries are the only inherent source of bargaining power. Second, the asymmetric information is two-sided so that neither partner has an inherent informational advantage. Third, while information asymmetries could affect bargaining in essentially any bargaining protocol, the Abreu and Gul (2000) model provides clear equilibrium predictions for the way in which asymmetric information should affect rational bargaining behavior, which is not a feature of other bargaining problems with incomplete information. In particular, for the simple version of the model that we implement (with only one irrational type), there is a unique equilibrium prediction: all subjects should mimic the demand of the only irrational type. This generates an interesting tension between the fair-division, which seems a natural norm in the absence of information asymmetries because of the symmetric bargaining protocol, and the more aggressive strategic posture, which is the unique equilibrium prediction. As shown by Embrey et al. (2015), many subjects appear to understand the strategic environment, with the majority either adopting the fair posture (an equal division of surplus) or mimicking the irrational type. Finally, the Abreu and Gul (2000) model identifies two additional features of bargaining, which seem important for understanding gender differences in real-world bargaining outcomes: (i) equilibrium bargaining delays generate inefficiencies, and (ii) even though the bargaining protocol is completely symmetric, information asymmetries generate asymmetric bargaining outcomes. Understanding gender differences in strategic posturing within the Abreu and Gul (2000) setting can therefore provide insights into the bargaining inefficiencies generated by men and women, and identify an alternative source of gender differences in bargaining outcomes that is related to information asymmetries rather than asymmetries of the bargaining protocol.

We therefore implement the basic design in Embrey et al. (2015) in two experimental conditions, which enables us to address two central research questions. First, in our control condition, subjects do not know the gender of their bargaining partner, allowing us to study whether there are gender differences in the Abreu and Gul (2000) bargaining framework when subjects are not able to make strategy choices contingent on their partner's gender. Second, in our treatment condition, genders are revealed, allowing us to study how knowledge of gender alters the strategic environment, and whether gender-revelation mitigates or exacerbates any gender differences in posturing behavior.

Related to our research questions, we report two main findings. First, in our control condition, we find significant gender gaps in posturing behavior: male subjects are much more likely to mimic the induced type, while female subjects are more likely to propose an equal division of the surplus. Our first finding therefore identifies gender differences in the Abreu and Gul (2000) framework. Second, we find that the treatment has a significant effect: revealing gender eliminates the gender gap in posturing behavior: if anything, female subjects are now more likely to mimic the induced type than male subjects, and less likely than male subjects to propose an equal division. We also observe some learning in the treatment over rounds. While in the control, the gender gaps in posturing behavior are consistent over the rounds, in the treatment meaningful gender gaps emerge only after subjects have gained some experience with the behavior of male and female opponents.

In addition to posturing behavior, we also look at the bargaining delays generated by subject-pairs who do not reach agreement in the first stage. We observe that male-male pairs generate significantly longer delays than either mixed-gender or female-female pairs, but only in the treatment condition where gender is revealed and only when one of the subjects in the pair has mimicked the induced type. Rather than achieving the desired concession, male subjects therefore generate significant bargaining inefficiencies when they adopt a strategic posture against a male opponent, suggesting that mimicking the induced type is not viewed as a credible signal of commitment.

Overall, our results indicate that male subjects may anticipate and/or learn that their strategic posture is not a credible signal, leading male subjects to adopt a strategic posture less often in a treatment condition where their gender is revealed. As such, our results provide new insights on gender differences in bargaining problems with asymmetric information. In particular, our findings indicate that revealing gender seems to alter the strategic environment in a meaningful way: parties accounting for gender mitigates differences in the way that men and women approach the bargaining problem.

The remaining paper is organized as follows. Section 2 discusses related literature. Section 3 presents the design, and Section 4 provides an overview of the data and describes our dependent variables. We present and interpret our main qualitative findings in Section 5. In Section 6, we complement the qualitative analysis with regression results. Section 7 concludes. An appendix provides additional data analysis and experimental instructions.

2 Related literature

Our experimental design is informed by the bargaining model in Abreu and Gul (2000). They study a bilateral, repeated-offer bargaining problem where, with some probability, agents can be irrational types who never deviate from a fixed bargaining position. Irrational types generate asymmetric information, which rational players can exploit to induce concessions from their partner. There are two key equilibrium predictions. First, rational players will always mimic irrational types in their initial demands, thereby deviating from an ostensibly fair allocation (such as the Nash bargaining solution). In the continuation, they follow a random stopping rule, thereby generating a war of attrition that can last as long as each party remains uncertain about their partner's type. As such, the second key prediction is that uncertainty about types generates bargaining delays. This prediction is in stark contrast to models without asymmetric information—such as the Rubinstein (1982) alternating-offer model—where agreement is reached immediately in equilibrium. As a result, uncertainty about types provides a theoretical explanation for bargaining outcomes that are inefficient and unequal.

While Abreu and Gul (2000) focus on a specific bargaining environment, they show that their equilibrium predictions are independent of key features of the bargaining protocol. Moreover, their model has become a benchmark for thinking about bargaining problems with asymmetric information, and the insights from their analysis have been applied to study many other features of bargaining behavior (Damiano et al., 2012; Fanning, 2016, 2018), as well as a broad range of other topics including political competition (Baron, 2003), banking (Povel, 2005) auctions (Kwiek, 2011), and search markets (Atakan and Ekmekci, 2014; Özyurt, 2015).

Given the wide-ranging influence of the Abreu and Gul (2000) model, Embrey et al. (2015) develop an experiment to test the model's key predictions. In their design, paired subjects first simultaneously propose how to divide a fixed pie and, if initial demands are not compatible, enter a second-stage continuous time concession game. This symmetric bargaining protocol has the advantage that neither partner has a first-mover advantage, and that there is a natural trade-off between a 50:50 norm (splitting the pie equally) and mimicking an irrational type. To induce uncertainty about types, subjects can be randomly matched with computers who are coded to demand a fixed bargaining position and never concede. Subjects know the strategy followed by computers, as well as the likelihood of a computer match, but never know whether they have been paired with a computer or a human subject. The key predictions of Abreu and Gul (2000) apply: in equilibrium, subjects should mimic the computer types in the first stage, generating a war of attrition that leads to costly delays in the second stage.

Embrey et al. (2015) conduct the experiment with a control condition, where there are no induced types, and five different treatments that vary the domain of demands in the first stage, as well as the proportion and initial demands of the computer players. The gender of subjects is never revealed in their experiment. Our control condition, where gender is also not revealed, is analogous to their treatment U_1 in which initial demands are unconstrained and there is a single induced type that demands two-thirds of the pie. All other treatments in Embrey et al. (2015) differ substantially from our control condition because there is always also an induced type who demands the 50:50 split.

Overall, Embrey et al. (2015) find broad support for the theoretical predictions of the Abreu and Gul (2000) model. For example, in their simplest treatment U_1—which we adapt for our design—they find that subjects demand two-thirds (mimicking the computer) a quarter of time, compared to one tenth in a control condition without computer players. Moreover, delays are significantly longer in all treatments with computer players relative to the control, and inducing irrational types therefore generates significant inefficiencies. The findings in Embrey et al. (2015) therefore suggest that inducing irrational types is effective at making the asymmetric information problem salient to subjects.

The qualitative patterns in our control condition are similar to those in Embrey et al. (2015), but with less variation in initial demands. While the majority of subjects in their treatment U_1

demand either an equal split or mimic the inducted type, there is still considerable heterogeneity. By contrast, more than 80% of the subjects in our experiment either demand an equal split or mimic the induced type. The pattern of initial demands we see in our data is therefore even more line with the equilibrium predictions of the Abreu and Gul (2000) framework, suggesting that many subjects in our experiment understand the asymmetric information bargaining problem.

It is not possible for us to identify why the behavior of subjects in our experiment follows theoretical predictions more closely than in Embrey et al. (2015), but there are differences in both the cultural context and the implementation of the design. First, our experiment was conducted in Magdeburg (Germany), while Embrey et al. (2015) conducted their experiment in New York City (USA). A literature following Roth et al. (1991), has shown that cultural context matters for bargaining behavior. While Roth et al. (1991) consider a series of ultimatum games, cultural context is also likely to matter in the Abreu and Gul (2000) framework, where behavior depends on a trade-off between fairness norms and strategic postures—as well as the credibility of signals—which may differ in Germany and the USA. Second, Embrey et al. (2015) generate uncertainty about types by randomly matching subjects with computer players who follow a fixed strategy, while we depart from their design by randomly constraining (human) subjects to follow a fixed strategy. Constraining human subjects to play the role of irrational types is important for our gender-revelation treatment, where subjects are made aware of their partners gender. However, there is also evidence that subjects behave differently when they know that they can be matched with a computer player (Farjam, 2019). Given these differences in the cultural context and design, the broad consistency of our findings suggests that the main conclusions in Embrey et al. (2015)—that the behavior of many subjects is line with equilibrium predictions—does replicate, but with some variation in the distribution over initial demands.

Gender differences in posturing behavior is not a focus of Embrey et al. (2015)'s study, but they do collect data on gender for two out of five sessions in each of their treatment conditions. For their treatment U_1, they collect information on gender for two sessions with 16 subjects, each of whom played the game 15 times (with random re-matching between rounds). This does not provide enough independent observations for a robust statistical analysis of gender differences in posturing behavior, especially given the greater variation in initial demands. However, qualitatively, the data from the two sessions with information on gender in treatment U_1 are consistent with the findings we report for our control condition in Section 5. In particular, in Embrey et al. (2015)'s study, female subjects demand an equal split more often than male subjects (33% of the time vs. 20% of the time), and male subjects mimic the induced type more often than female subjects (45% of the time vs. 15% of the time).

We build on the design in Embrey et al. (2015) to study gender differences in posturing behavior for the Abreu and Gul (2000) framework. To address our main research questions, we adapt their treatment (U_1) in two alternative conditions, which differ in terms of whether gender is revealed (treatment) or not (control). The control allows us to identify gender differences when subjects are not able to condition strategic behavior on their partner's gender, while comparing the control with the treatment shows how these gender differences are mitigated or exacerbated when subjects can condition bargaining strategies on gender.

Our findings contribute to a growing literature on gender differences in bargaining. Empirical studies provide some evidence that men are more aggressive—and more successful—in wage and price negotiations, but the findings are context-dependent. For instance, Babcock and Laschever (2009) find that 57% of male graduate business students negotiate their starting salaries compared to 7% of women, and that male starting salaries are 7.6% higher. In the market for new cars—where price negotiation is not unusual—Ayres and Siegelman (1995) find men pay lower prices based on data from tester audits. However, Goldberg (1996) finds no gender difference in prices paid for new cars using consumer expenditure survey data, and Harless and Hoffer (2002) also find no gender difference using transaction price data. In the housing market, Harding et al. (2003) find some evidence that women have less bargaining power, but evidence on gender differences in bargaining skills using data from real estate agents is inconclusive (Seagraves and Gallimore, 2013).

The challenge of empirically isolating bargaining behavior has also motivated an experimental literature. Eckel and Grossman (2001) and Solnick (2001) study differences in the bargaining behavior of men and women in an ultimatum game. Both studies find that men and women make similar offers as the proposer, but that offers to male responders are higher than offers to female responders. Eckel and Grossman (2001) find that women accept lower offers, while Solnick (2001) find the opposite. As a result, Eckel and Grossman (2001) find women receive higher earnings, whereas Solnick (2001) find that men earn more.[4]

In our framework, we find that men demand more than women when gender is not revealed, but that these gender differences disappear with gender-revelation. The influence of revealing gender has also been observed in other bargaining problems. For instance, in ultimatum games, Eckel and Grossman (2001) find that a given offer is more likely to be accepted when it is known to come from a women, a finding they relate to chivalry, and that female-female pairs almost always end with accepted offers, a finding they interpret as solidarity; in an alternating-offer wage bargaining experiment, Dittrich et al. (2014) find that male employers pay lower wages to female employees than female employers pay to male employees; in a two-stage power-to-take game, Sutter et al. (2009) find that subjects take more in the first stage and destroy more in the second stage in same-gender pairs than in mixed-gender pairs; and, using data from a

[4]There are several differences in the experimental designs: (i) while Solnick (2001) consider one-shot games, the game in Eckel and Grossman (2001) is repeated allowing for learning in later rounds; (ii) Solnick (2001) uses the strategy method for responders, which Eckel and Grossman (2001) argue may lead to more rejections because subjects fail to fully understand the simultaneous nature of the game; and (iii) Eckel and Grossman (2001) study face-to-face interactions, while Solnick (2001) uses first names to reveal gender.

Spanish TV show with bilateral bargaining situations, Hernandez-Arenaz and Iriberri (2018) find women ask for a lower share of a monetary amount against men than against other women.

The novelty in our findings for delays is that the effect of partner's gender primarily arises when the partner has adopted a strategic posture. In particular, we find that male-male pairs experience higher delays than other gender-pairs, suggestive of a rivalry between men. However, this gender effect is only present when one of the partners has mimicked the induced type, and therefore appears related to the way that subjects respond to strategic postures. This strategic dimension of the gender effect can be rationalized when subjects have different expectations about the willingness of men and women to feign commitment. Based on the previous literature, such priors could reflect a gender stereotype that women are less aggressive than men (see, e.g. Kray and Thompson, 2004; Bowles, 2014). For instance, in a survey of dictator and ultimatum games, Eckel et al. (2008) find that—while there are gender difference in behavior—these differences are small compared to the difference in what men and women are expected to do, suggesting that stereotypes can have a significant impact on negotiations.

Alternatively, different expectations about men and women could also be rationalized by intrinsic gender differences, consistent with the gender differences we observe in our control condition. Such gender differences could reflect a variety of characteristics (e.g., selfishness (Eckel and Grossman, 1998), willingness to deceive (Dreber and Johannesson, 2008), or competitiveness (Niederle and Vesterlund, 2007)), which have been documented to differ for men and women in the prior literature.[5] As such, the treatment effect for delays we observe in our setting are consistent with a form of statistical discrimination, whereby subjects make inferences about their partner's strategic intent (which is unobservable) from their partner's gender (which is observable in our treatment).[6]

3 Experimental design

The experiment consists of two parts and employs a between subject design. Part 1 sets up the gender-revelation; part 2 is a bilateral bargaining problem. There are two conditions that differ in whether partner's gender is revealed (treatment) or not (control). We first give an overview

[5]In a review of the literature, Niederle (2014) argues for a more nuanced view regarding gender differences in behavior. In particular, while results on gender differences in competition seem robust, the findings that women are more altruistic seem sensitive to details of the experimental designs. Overall, while the literature has found systematic gender differences in behavior, it is not clear if these reflect fundamental differences in intrinsic characteristics, or context-dependent responses to decision-making problems.

[6]There is evidence from the prior literature that statistical discrimination is relevant in bargaining environments. For instance, in a field experiment on the taxi market in Peru, Castillo et al. (2013) find women are quoted lower prices than men—indicating that men are seen as having higher valuations—but this gender difference disappears when people can send a signal of valuation before arranging a fare, indicating that the initial gender difference is due to statistical discrimination.

of the experimental design, and then discuss the gender-revelation in more detail. Instructions and screenshots, translated from German, are provided in Appendix A.2.

In each experimental session, 16 subjects are randomly assigned to separate booths and given a sealed envelope. The instructions for part 1 are then read aloud. Subjects are asked to open their envelopes to find a unique pseudonym on a slip of paper. The pseudonym takes the form "player [City]" where city is the capital of a European country (e.g., Amsterdam, Copenhagen, or Oslo). In the control, subjects are asked to type the pseudonym into a box on their computer screens. In the treatment, subjects are asked to put on headsets and say the pseudonym into a microphone to record it as an audio file. Once part 1 of the experiment is concluded, the instructions to part 2 are handed out and read aloud.

Part 2 of the experiment has 15 rounds. In each round, subjects are paired using perfect stranger matching, and the pairs engage in a two-stage bilateral bargaining game. Before stage 1, the pseudonym of the partners is revealed. In the control, subjects see a screen for 15 seconds on which the pseudonym is displayed (not revealing gender). In the treatment, subjects see a blank screen and hear the pseudonym via the recording from part 1 (revealing gender). Gender-revelation is the only difference between the two conditions.

In the first stage of the bargaining game, subjects simultaneously demand a share out of 30 points for themselves. If the sum of the two demands is less than or equal to 30, then the demands are compatible and the round ends. Subjects then receive their demands, with any remaining points split equally. If the sum of the two demands exceeds 30, then the demands are not compatible and the pair proceeds to the second stage.

The second stage is a continuous-time concession game. Each second t the demanded share from the first stage is discounted by $\exp(-0.001t)$. Either subject can end the game at anytime by pressing a concession button. The partner who does not concede receives their discounted demand. The partner who concedes receives the amount left over, i.e., the total points after discounting and subtracting the discounted demand of the non-conceding partner. To aid the bargaining pairs in the second stage, a 2×2 matrix displays both the subject's and their partner's payoffs, discounted in real-time, for the scenario where the subject concedes and the scenario where the partner concedes. When the second stage of the experiment is concluded, each subject is shown their payoff from the round and is then randomly assigned a new partner.

Once all 15 rounds are complete, subjects receive their payoffs. As in Embrey et al. (2015), we use the lottery method to induce risk neutrality (Roth and Malouf, 1979). Specifically, the payoffs are provided as the outcome of a lottery in which the probability of winning 20 euros is determined by the points subjects received in each round. An additional show up fee of 10 euros is paid to all subjects. Inducing risk neutrality has the advantage that we can interpret our results in terms of gender differences in the willingness to engage and commit to strategic postures, free from the confounds of differences in risk attitudes. This feature is important because the literature has found gender differences in risk attitudes, and this could affect bargaining behavior in a design with monetary payoffs (Croson and Gneezy, 2009).[7]

In the experiment, subjects can be one of two types, spade or diamond. Subjects are informed of their type at the start of part 2, but never learn the type of any of their partners. Types are fixed throughout the 15 rounds. Diamonds are free to play the game as they wish. Spades are our *induced types*, who are constrained to follow a fixed strategy: they demand 20 in the first stage and cannot concede in the second stage. Out of the 16 subjects, 14 are diamonds and 2 are spades. While the type is private information, all subjects are informed about what each type is required to do and the proportion of each type in the session.

Gender-revelation. An important aspect of our design is that subjects in the treatment condition receive a signal that conveys information about their partner's gender. The most direct way to reveal gender would be to tell subjects their partner's gender explicitly. However, this approach could induce experimenter demand effects, which may be especially pronounced in our setting because subjects are matched 15 times and being informed about their partner's gender each time may then affect behavior. One could also provide subjects with the first name of the partner, but this provides information to the subjects that is commonly kept confidential for reasons of anonymity.

An alternative approach used by Coffman (2014) provides subjects with pictures of their partners. However, Bordalo et al. (2019) point out that pictures potentially reveal other attributes such as age, race, or attractiveness, which Eckel and Grossman (2001) show can also affect bargaining behavior. Bordalo et al. (2019) therefore reveal gender by giving subjects a brief opportunity to hear their partner's voice. While there is no perfect signal of gender that conveys no other information at all, hearing the voice arguably reveals gender without much else. We implement this approach using the brief recording of pseudonyms. We are cognizant that hearing the voice of the partner may also reveal additional information and, like Bordalo et al. (2019), we therefore limit the gender-revelation to a short voice-snippet, which is long enough to infer gender but makes it difficult to infer confounds such as "friendliness".[8] In a post-experiment questionnaire subjects were asked how often they recognized the gender of the partner, and 91.67% of subjects reported recognizing gender in at least 14 out of 15 rounds.

[7]There is some evidence from psychology and economics that women are more risk-averse than men (see, e.g., Byrnes et al., 1999), but results are mixed. In particular, whether there are gender differences can depend on the risk elicitation method used (see, e.g., Niederle, 2014), and may also depend on the context (see, e.g., Pondorfer et al., 2016).

[8]Bordalo et al. (2019) use an attendance check where subjects say the word "here" to reveal gender. We use the pseudonym because we have rematching of subjects over 15 rounds.

4 Data

In this section, we first provide an overview of the experimental sessions. We then describe the dependent variables in our analysis, followed by the main independent variables.

4.1 Overview of experimental sessions

The experiment was conducted at the MaxLab in Magdeburg (Germany), and was coded using z-Tree (Fischbacher, 2007). Recruitment was carried out using hroot (Bock et al., 2014). In total, 160 subjects participated in 10 sessions—4 control sessions and 6 treatment sessions—and each subject played 15 rounds in their session, resulting in total of 2400 observations. In every round, two subjects were randomly selected to be induced types. Although each session could accommodate a maximum of 16 subjects, we invited 20 subjects to compensate for no-shows.

Subjects were randomly assigned to the control and treatment conditions: 64 (35 male and 29 female) were assigned to the control, and 96 (48 male and 48 female) were assigned to the treatment. To promote gender-balance, an equal number of female and male subjects were invited to each session. Three sessions achieved gender-balance (2 in the treatment and 1 in the control), while the percentage of male participants in the remaining sessions ranged from 43.75% to 62.5%.[9] Since subjects were randomly assigned to the role of induced type, it was not possible to achieve an equal number of male and female induced types in each session. In particular, we had two sessions in which both induced types were female (one control and one treatment), two sessions in which both were male (one control and one treatment), and six sessions in which one induced type was male and the other was female.

We also collected data on major and age using a post-experiment survey. Based on the survey, approximately 24% of subjects were economics or business majors. The age range was between 19–35 years old, with a median age of 24. Average earnings from the experiment were €17.88, and the average session time was $\approx$ 90 minutes.

4.2 Dependent variables

In stage 1 of each round, subjects simultaneously demand a share of the points. We use the demand data from stage 1 to look at how the propensity for strategic posturing is related to gender and the gender-revelation treatment. First stage behavior is the outcome of a relatively straightforward simultaneous decision by subjects—whether to adopt a fair or strategic posture—and we therefore analyze the stage 1 data at an individual level.

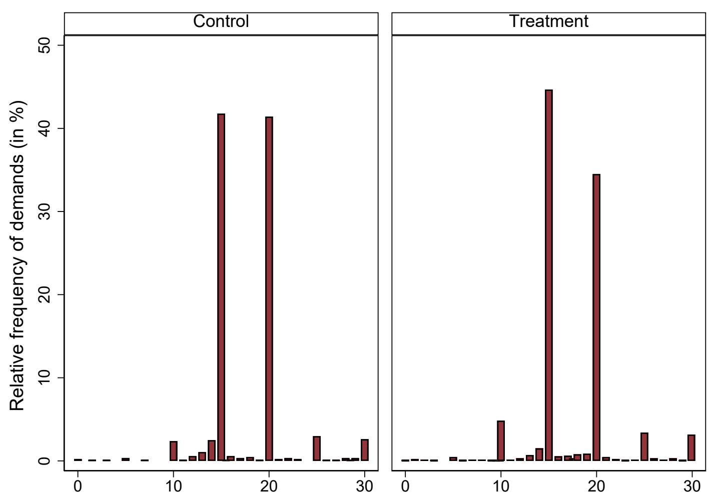
Our main research questions are concerned with how gender and gender-revelation impact

[9] In Appendix A.1.1, we show that our regression results in Section 6 are robust to controlling for whether the session is gender-balanced or not.

posturing behavior, and are therefore only concerned with data from stage 1. However, bargaining delays—which generate inefficiency—are an important and interesting prediction of the Abreu and Gul (2000) model, which is directly related to the two-sided asymmetric information problem. We therefore also use data from stage 2, where subjects face a continuous-time concession game, to look at how bargaining delays are related to gender and the gender-revelation treatment. Since a single delay is observed for each pair in each round, we analyze the stage 2 delay data at a pair-level.

Stage 1: posturing behavior. Our first set of dependent variables measures posturing behavior in stage 1 in terms of the demands of the subjects. Figure 1 shows the histogram of demands in the control and treatment, excluding demands from subjects in the role of induced types (sample size = 2100, see Table 1 for details).

Figure 1: Histogram of demands in stage 1 in control and treatment



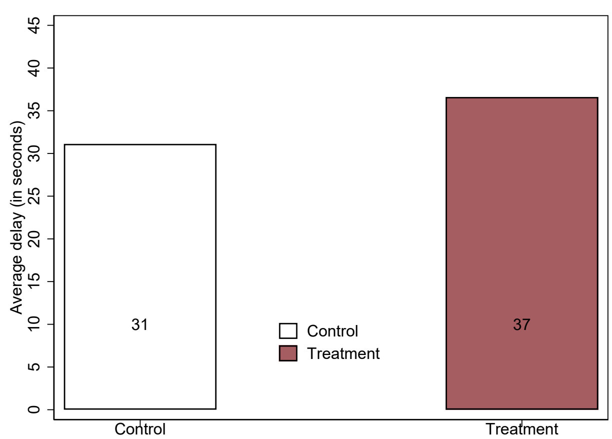
While there are some deviations, the vast majority of demands (more than 80% in both the control and treatment) are either 15 or 20. Instead of analyzing the demand directly as a continuous variable, we focus on whether a subject demands 15 or 20 by coding three binary variables: $demand_{ir}^{15}$, which takes value 1 if subject i demanded 15 in round r and 0 otherwise; $demand_{ir}^{20}$, which takes value 1 if subject i demanded 20 in round r and 0 otherwise; and $demand_{ir}^{other}$, which takes value 1 if subject i made a demand other than 15 or 20 in round r. A subject demanding 15 is proposing a 50-50 split, and is therefore adopting a *fair posture*; a subject demanding 20 is adopting a *strategic posture*, which mimics the induced type. For the analysis of posturing behavior in Sections 5 and 6, we focus on demands of 15 and 20 because of their direct economic interpretation.[10] Alternatively, one could also consider the

[10]In Appendix A.1.3, we also provide results from a multinomial regression with demand of 15, demand of

initial demands directly as a continuous dependent variable. However, while an increase in the initial demand from 20 to 21 represents a more aggressive demand in principle, it also represents a departure from the behavior of the induced type and is therefore a potential deviation from the equilibrium prediction. As such, marginal effects for such a demand are difficult to interpret, and we prefer to focus on the binary variables $demand^{15}$ and $demand^{20}$, which constitute the vast majority of our observations and have clearer economic interpretations.

Stage 2: bargaining delays. We focus our analysis of stage 2 on the delays that arise for bargaining pairs who do not reach agreement in stage 1. When subjects i and j are matched in a round, they form a pair ij. If the pair ij do not reach agreement in stage 1 (i.e., the sum of initial demands for i and j exceeds 30), they enter stage 2 and we observe $Delay_{ij}$, which measures the time in takes in seconds until one of the subjects concedes. All pairs in our data reached agreement in stage 2. Figure 2 illustrates the average delays in the control and treatment, excluding pairs that did not reach stage 2 or where at least one partner was an induced type (sample size = 627 observations, see Table 2 for details). As we can see, average delays are slightly longer in the treatment than in the control.

Figure 2: Average delays in control and treatment



Analyzing second stage behavior is more complex because subjects engage in a continuous-time bargaining process in which they continuously decide whether to concede or not based on what they and their partner demanded in stage 1, while updating beliefs about whether their partner will concede first. While the variable $Delay_{ij}$ does not capture the full spectrum of second stage behavior, it provides a simple measure of two important feature of the bargaining environment. First, bargaining delay is a key measure of the inefficiency generated by the bargaining process in the Abreu and Gul (2000) setting. Second, delays provide a simple

20, and demand other than 15 or 20 as three categories.

measure of a subject's success with a strategic posture: if the signal of being an induced type is credible, it is optimal for the partner to concede to avoid further delay. Therefore, if a demand of 20 is a credible signal, we would expect to observe shorter delays in pairs where one of the partners adopted a strategic posture in stage 1 than in pairs where neither partner adopted the strategic posture. We therefore also analyze two subsamples: (i) the 527 pairs where at least one of the subjects adopted the strategic posture in stage 1, and (ii) the remaining 100 pairs where neither partner adopted the strategic posture in stage 1.[11]

4.3 Independent variables

While there are systematic patterns in the aggregate demand and delay data, there is also considerable variation. Our aim is to identify whether some of this variation can be explained by systematic difference across genders and/or differences across the control and treatment.

We identify treatment effects by the random assignment of subjects to the control and treatment conditions. For individual-level data in stage 1, $treatment_i$ is a treatment dummy that takes value 1 if subject i is in the treatment condition and 0 otherwise. For the paired-data in stage 2, $Treatment_{ij}$ is a treatment dummy that takes value 1 if the pair ij is in the treatment condition and 0 otherwise.

We use gender dummies to study gender effects. For individual-level data, $male_i$ is a gender dummy that takes value 1 if subject i is male and 0 otherwise. For pair-level data, $Male_{ij}$ is a gender composition dummy that takes value 1 if subjects i and j are both male, and 0 otherwise; $Female_{ij}$ is a dummy variable taking value 1 if subjects i and j are both female, and 0 otherwise; and $Mixed_{ij}$ is a dummy taking value 1 if one of the subjects in pair ij is male and the other is female, and 0 otherwise. The gender-composition dummies are perfectly collinear, and we treat $Male_{ij}$ as the omitted variable for the regression analysis in Section 6.

5 Results

In this section, we first address our main research questions: (i) we look at demands in the control to ask whether there are gender differences in posturing behavior when subjects cannot condition strategic behavior on gender, and (ii) we ask how knowledge of gender impacts posturing behavior by comparing demands in the control with the treatment. Second, we look at

[11] Subsample (i) could be further divided into pairs where (ia) exactly one subject demands 20, or (ib) both subjects demand 20. In Appendix A.1.4, we show that the distribution of delays is essentially the same for (ia) and (ib), and we merge these categories in order to increase the power for the statistical analysis. Subsample (ii) necessarily contains only pairs where at least one partner demands something other than 15 or 20 (since neither demand 20 and the pair does not reach stage 2 if both demand 15). While demands of 20 and 15 have natural interpretations as strategic or fair postures, it is more difficult to rationalize demands other than 15 or 20, and we look at subsample (ii) only for comparison.

delays in stage 2 to analyze how bargaining inefficiencies depend on the gender-composition of the pair, and shed light on the credibility of strategic postures.

For the demand analysis, we look at conditional means for adopting the fair posture ($demand^{15}$) and strategic posture ($demand^{20}$), broken down by gender (male or female) and condition (control or treatment), with corresponding t-tests for mean-comparisons.[12] We also look at posturing behavior round-by-round for evidence of potential learning effects. For the delay analysis, we look at conditional means, broken down by gender composition of the pair (male-male, mixed-gender, or female-female) and condition (control or treatment), with corresponding t-test for mean comparisons. In Section 6, we complement the descriptive analysis in this section with probit regressions for the the binary demand variables in stage 1 and OLS regressions for the continuous delay variable in stage 2.

5.1 Posturing behavior

Table 1 summarizes the demand data from stage 1. For the demand analysis, we exclude subjects who were in the role of induced type, which removes 300 observations ($2 \times 15 = 30$ observations for each of the 10 sessions) from the total number of observations ($16 \times 15 = 240$ for each of the 10 sessions), leaving a sample size of 2100 observations.

Table 1: Demands in stage 1

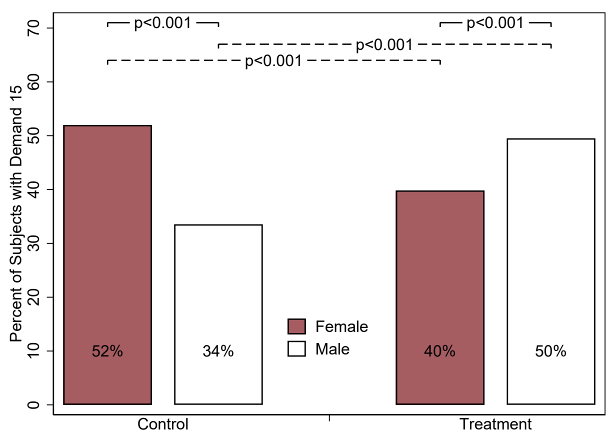
	Panel A: Control						
	(1) Male			**(2) Female**			**(3) Total**
	Mean	**Std**	**Obs**	**Mean**	**Std**	**Obs**	
(1) Demand 15	33.548	47.267	156	52.000	50.027	195	351
(2) Demand 20	51.828	50.020	241	28.533	45.218	107	348
(3) Demand Other	14.624	35.372	68	19.467	39.647	73	141
(4) No. of Control Obs.			*465*			*375*	*840*
	Panel B: Treatment						
	(1) Male			**(2) Female**			**(3) Total**
	Mean	**Std**	**Obs**	**Mean**	**Std**	**Obs**	
(5) Demand 15	49.524	50.037	312	39.841	48.996	251	563
(6) Demand 20	31.429	46.460	198	37.619	48.481	237	435
(7) Demand Other	19.048	39.299	120	22.540	41.818	142	262
(8) No. of Treatment Obs.			*630*			*630*	*1260*
(9) Total			***1095***			***1005***	***2100***

[12]In the treatment, a subject's demand in stage 1 could also depend on their partner's gender. To assess whether there are partner-gender effects in the treatment condition, we also estimate a probit regression for $demand^{15}$ (respectively, $demand^{20}$) on $male_i$, $male_j$, and $male_i * male_j$, where $male_i$ is the gender dummy of the subject and $male_j$ is the gender dummy of the subject's partner. We do not find that partner's gender has a meaningful impact on a subject's demand and therefore defer these regression results to Appendix A.1.2.

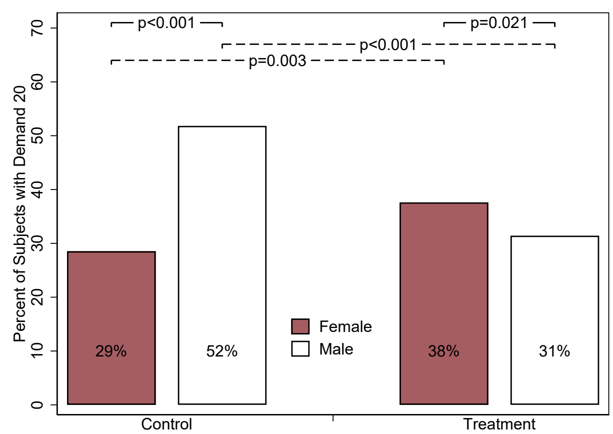
For the control, Panel A provides the number of observations for demand of 15 (Row 1), demand of 20 (Row 2), and demand other than 15 or 20 (Row 3), first for male subjects (Column 1) and then for female subjects (Column 2). Each cell also provides the mean (i.e., relative frequency in % that the demand dummy takes value 1), along with the standard deviation. Panel B provides the analogous summary statistics for the treatment.

We perform two sets of hypotheses tests for the fair (demand of 15) and strategic (demand of 20) postures. First, to assess whether there is a gender gap in the control (respectively, treatment), we test the null hypothesis that the conditional mean for male and female subjects is equal in the control (respectively, treatment). Second, to assess whether there is a treatment effect for male (respectively, female) subjects, we test the null hypothesis that the conditional mean for male (respectively, female) subjects is equal in the control and treatment.

Figure 3: Demands in stage 1



(a) Demand 15



(b) Demand 20

Note: Subfigure (a) displays the percentage of subjects demanding 15 in stage 1 by gender (male or female) and condition (control or treatment). Subfigure (b) displays the percentage of subjects demanding 20 in stage 1 by gender (male or female) and condition (control or treatment). Conditional means are based on the observations summarized in Table 1, rounded to the nearest integer. Top horizontal bars show p-values for a t-test of the null hypothesis that the means are equal.

Figure 3 summarizes the information from Table 1 by illustrating the conditional means for the fair and strategic posture. Figure 3(a) shows the conditional mean of $demand^{15}$ for both genders in the control condition (first two bars) and treatment condition (last two bars), while Figure 3(b) shows the conditional mean of $demand^{20}$ in the control and treatment conditions. We also provide p-values for the two-sided t-tests of the null hypothesis that the corresponding conditional means are equal. The test statistics are computed from the conditional means, number of observations, and standard deviations in Table 1.

Gender gap in the control. The first two bars in Figure 3(a) and corresponding p-values show that there is a significant gender gap for the mean of *demand*[15] in the control condition: female subjects adopt the fair posture 52% of the time, while male subjects adopt the fair posture only 34% of the time. The first two bars in Figure 3(b) show that, for demand of 20, the gender gap goes in the opposite direction: in the control, male subjects adopt the strategic posture 52% of the time while female subjects do so only 29% of the time.

We therefore find a clear answer to the first research question: there are significant gender differences in our setting. When it is not possible to condition strategies on gender, men are more likely to aggressively pursue their own self-interests by adopting a strategic posture, while women are more accommodating in seeking a fair division of the surplus.

Treatment effects. Our second research question addresses how gender-revelation impacts posturing behavior. For this, we look at the treatment effects from gender-revelation.

For male subjects, we see a significant treatment effect for demand of 20: while male subjects adopt a strategic posture 52% of the time in the control, they do so only 31% of the time in the treatment. Male subjects also demand 15 less often in the control (34%) than in the treatment (50%), indicating that male subjects are less aggressive in pursuing their self-interests when gender is revealed. For female subjects, we see the opposite pattern: while female subjects adopt the strategic posture only 29% of the time in the control, they do so more often in the treatment (38% of the time). Female subjects also adopt the fair posture more often in the control (50%) than in the treatment (40%).

As a result, the significant gender gaps we observe for fair and strategic postures in the control condition reverse when gender is revealed: in the treatment condition, female subjects demand 15 less often than male subjects (40% vs. 50%), and demand 20 more often than male subjects (38% vs. 31%).

For the mean-comparison tests in Figure 3, all of these treatment effects are significant, but only the statistical significance of the treatment effect for male subjects demanding 20 is robust when we cluster standard errors at the session level (see Section 6). However, the regression analysis shows a significant, and large treatment effect on the gender gaps for demand of 15 and demand of 20. Female subjects adopt the fair posture significantly more often than male subjects in the control, but this gender gap disappears in the treatment. On the other hand, male subjects adopt the strategic posture significantly more often than female subjects in the control, but this gender gap also disappears in the treatment.

The answer to our second research question is therefore also clear: gender-revelation has a significant impact on posturing behavior, eliminating the significant gender differences observed in the control where subjects cannot condition bargaining strategies on their partner's gender.

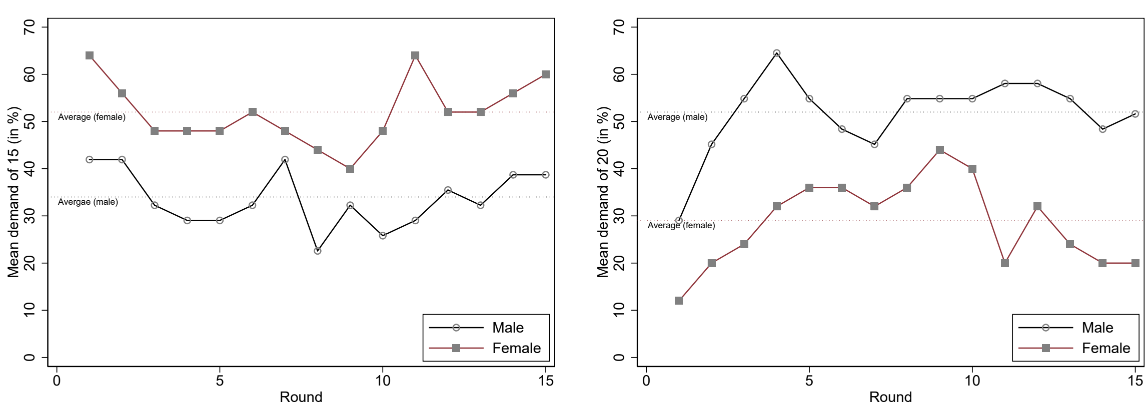
Learning. Finally, we also look at the patterns of fair and strategic postures round-by-round in the control and treatment conditions.

Figure 4 illustrates the relative frequencies for the fair and strategic posture by male and female subjects for each round in the control. Figure 4(a) shows that, while the relative frequency of demands fluctuates slightly over the rounds, female subjects consistently demand 15 more often than male subjects in the control, while Figure 4(b) shows that male subjects consistently demand 20 more often.

However, in the treatment we do see some evidence that the experience subjects gain over initial rounds impacts behavior. Figure 5(a) illustrates the mean demand of 15 for male and female subjects in the treatment round-by-round. We see that male subjects generally adopt the fair posture more often than female subjects, but a meaningful gender gap only emerges after the first 10 periods. Figure 5(b) illustrates a similar pattern for the strategic posture: female subjects generally demand 20 more often than male subjects, but a noticeable gender gap only emerges after subjects gain experience.

All subjects (across all our treatment sessions) encounter both male and female partners in the first 10 rounds and, in the treatment condition, therefore learn how the behavior of subjects is related to gender. We therefore observe patters that suggest some gender-related learning effects as subjects gain experience with the bargaining behaviors of male and female opponents.

Figure 4: Round-by-round demands in control

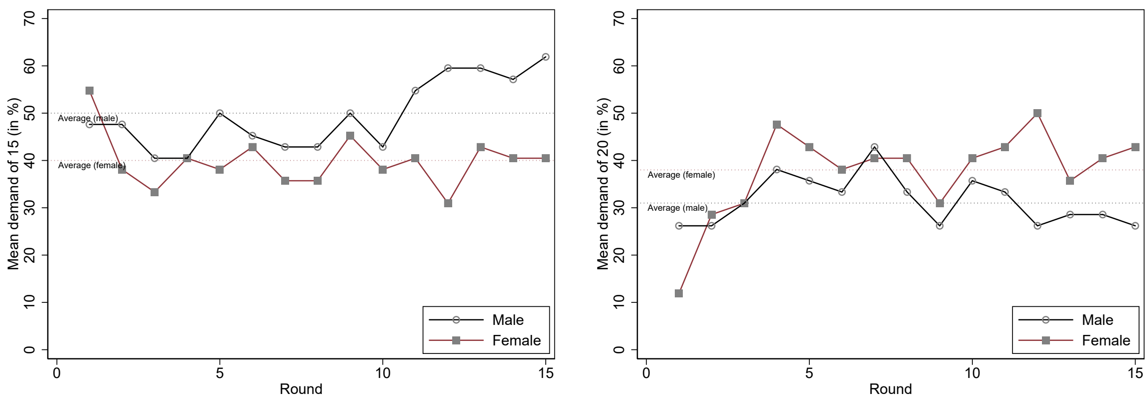


(a) Demand 15 in control

(b) Demand 20 in control

Note: Subfigure (a) displays the percentage of observations with demand of 15 by male and female subjects in each round of the control. Subfigure (b) displays the percentage of observations with demand of 20 by male and female subjects in each round of the control. Average over all periods for male and female subjects are provided for comparison.

Figure 5: Round-by-round demands in treatment



(a) Demand 15 in treatment

(b) Demand 20 in treatment

Note: Subfigure (a) displays the percentage of observations with demand of 15 by male and female subjects in each round of the treatment. Subfigure (b) displays the percentage of observations with demand of 20 by male and female subjects in each round of the treatment. Average over all periods for male and female subjects are provided for comparison.

5.2 Bargaining delays

In the Abreu and Gul (2000) setting, two-sided asymmetric information generates inefficient equilibrium delays. We use second stage data to analyze how bargaining delays in our experiment are related to the gender-composition of the pair.

Table 2 summarizes the delay data from stage 2, broken down by gender-composition of the pair (male-male, male-female, or female-female) and condition (control or treatment). In total, there are 1200 pairs across the 15 rounds and 10 sessions. For the delay analysis, we first remove all pairs who did not reach stage 2 (303 pair observations). Of the remaining pairs, we also exclude pairs where one or both partners were an induced type (a further 270 pair observations). This leaves a sample size of 627 pairs.

Panel *A* provides the number of observations, average delays, and standard deviations for each of the gender-pairings in the control condition (Row 1) and the treatment condition (Row 2). Panel *B* provides the same summary statistics but for two subsamples, depending on the posturing behavior in stage 1. Rows 3 and 4 provide the number of observations, average delays, and standard deviations for each of the gender-pairings where at least one partner demanded 20 in stage 1. Rows 5 and 6 provide the number of observations, average delays, and standard deviations for each of the gender-pairings where neither partner demanded 20 in stage 1.

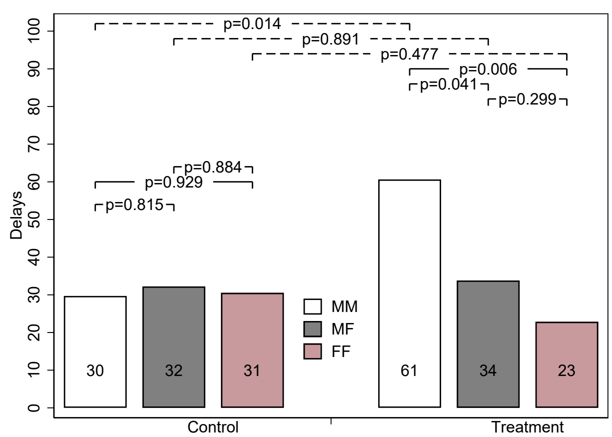
We perform hypotheses tests to assess whether the mean delay depends on the gender-composition of the pair, first in the control and then in the treatment. The null hypothesis for each test is that the conditional means are equal. We also test for treatment effects, where the

Table 2: Delay in stage 2

	Panel A: Delays									
Delay (secs)	**(1) Male**			**(2) Female**			**(3) Mixed**			**(4) Total**
	Mean	**Std**	**Obs**	**Mean**	**Std**	**Obs**	**Mean**	**Std**	**Obs**	
(1) Control	29.709	52.728	86	30.532	50.099	47	32.254	104.710	130	263
(2) Treatment	60.662	95.854	74	22.852	74.728	88	33.822	97.947	202	364
Total			*160*			*135*				*627*
	Panel B: Delay with Demands									
Delay (secs) -w demand 20	**(1) Male**			**(2) Female**			**(3) Mixed**			**(4) Total**
	Mean	**Std**	**Obs**	**Mean**	**Std**	**Obs**	**Mean**	**Std**	**Obs**	
(3) Control	30.630	54.163	81	26.594	38.940	32.000	34.931	110.524	116	229
(4) Treatment	67.123	100.228	65	24.155	81.595	71.000	24.488	52.063	162	298
Total			*146*			*103*			*278*	*527*
Delay (secs) -w no demand 20	**(1) Male**			**(2) Female**			**(3) Mixed**			**(4) Total**
	Mean	**Std**	**Obs**	**Mean**	**Std**	**Obs**	**Mean**	**Std**	**Obs**	
(5) Control	14.800	10.640	5	38.933	69.126	15	10.071	11.926	14	34
(6) Treatment	14.000	26.206	9	17.412	34.612	17	71.625	190.858	40	66
Total			*14*			*32*			*54*	*100*

null hypothesis is that the average delay for male-male (respectively, male-female or female-female) pairs is equal in the control and treatment.[13]

Figure 6: Delays in stage 2



Note: Average delays in stage 2 by gender composition of the pairs (MM = male-male, MF = mixed-gender, and FF = female-female). Left three bars display average delays in the control. Right three bars display average delays in the treatment. Conditional means are based on the observations summarized in Panel A of Table 2, rounded to the nearest integer. Top horizontal bars show p-values for a t-test of the null hypothesis that the means are equal.

For the control and treatment, Figure 6 illustrates the average delays for male-male pairs

[13]We also present p-values for treatment effects by gender-composition. The null hypothesis for these tests is that the conditional mean delay for male-male (respectively, male-female or female-female pairs) in the control and treatment are equal. None of these treatment effects is significant.

(MM), mixed-gender pairs (MF), and female-female pairs (FF). The figure also provides p-values for the two-sided t-tests of the null hypothesis that the corresponding conditional means are equal. The test statistics are computed from the conditional means, number of observations, and standard deviations in Table 2.

In the control, where subjects are not aware of their partner's gender, the first three bars show that there are no significant differences between the delays experienced by these three pair compositions. However, in the treatment, where partner's gender is revealed, the average delays are affected by gender composition. In particular, average delays are considerably longer for male-male pairs (61 seconds), than for pairs where one or both subjects are female (34 and 23 seconds, respectively). Therefore, the bargaining inefficiencies generated when male-male pairs do not reach agreement in the first stage are significantly larger than for pairs where at least one of the subjects is female, but only when the subjects are aware of their partner's gender.

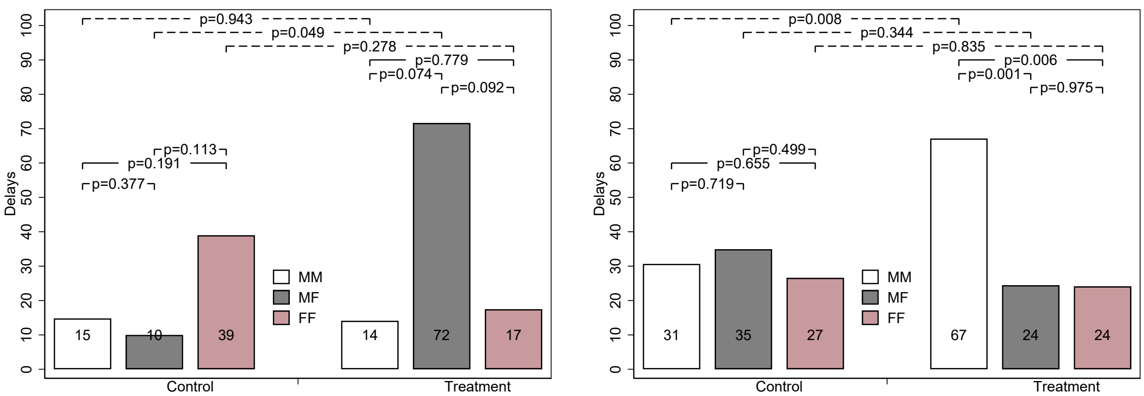
One potential reason for male-male pairs to experience longer delays in the treatment is that mimicking an induced type may not provide a credible signal of commitment if—as we observe in our stage 1 data for the control condition—male subjects are more likely to adopt strategic postures than female subjects. If longer delays are related to the credibility of strategic postures, we would expect to find that the difference in mean delays is more pronounced in pairs where at least one of the subjects has adopted a strategic posture in stage 1, than in pairs where neither partner mimicked the induced type.

In Figure 7, we therefore also look at the delays in the control and treatment when we distinguish between pairs in which neither subject adopted a strategic posture and pairs in which at least one of the partners adopted a strategic posture. The conditional means and p-values are based on the delay data summarized in Panel B of Table 2.

Figure 7(a) shows that, both in the control and treatment, there are no statistically significant differences between the delays for male-male, male-female and female-female pairs when neither partner adopted a strategic posture in the first stage. Since there are relatively few pairs who do not reach agreement in stage 1 when neither partner demanded 20, there are too few observations to reach strong conclusions. However, if anything, male-male pairs experience the shortest delays when neither partner is mimicking the induced type.

Figure 7(b) illustrates the delays for the subsample where at least one of the partners demanded 20 in stage 1. Again, we find no significant differences between male-male, male-female and female-female pairs in the control. However, in the treatment condition, where subjects can condition bargaining behavior on their partner's gender, male-male pairs experience significantly longer delays: the average bargaining delay for male-male pairs is 67 seconds compared to 24 seconds for both male-female and female-female pairs.

Figure 7: Delays in stage 2 conditional on posturing in stage 1



(a) Neither demands 20

(b) At least one demands 20

Note: Subfigure (a) displays the average delays in stage 2 by gender composition of the pairs when neither subject demanded 20 in stage 1. Subfigure (b) displays the average delays in stage 2 by gender composition of the pairs when at least one subject demanded 20 in stage 1. Conditional means are based on the observations summarized in Panel B of Table 2, rounded to the nearest integer. Top horizontal bars show p-values for a t-test of the null hypothesis that the means are equal.

Figure 7 therefore indicates that the significantly longer delays for male-male pairs in the treatment arise only when at least one subject adopted a strategic posture. This finding is consistent with the idea that a strategic posture is a less credible signal of commitment for male subjects than for female subjects, especially against a male opponent.

5.3 Summary of bargaining behavior and outcomes

In our control, gender is not revealed, and subjects therefore cannot condition strategic bargaining behavior on their partner's gender. In this control condition, we find significant gender effects for strategic posturing: male subjects are more likely to adopt a strategic posture than female subjects, while female subjects are more like to adopt a fair posture. Since subjects cannot condition behavior on their partner's gender, these gender differences could reflect characteristics such as selfishness, willingness to deceive, or competitiveness, which have been found to vary by gender in the previous literature (see, e.g., Eckel and Grossman, 1998; Dreber and Johannesson, 2008; Niederle and Vesterlund, 2007). However, outside the laboratory, bargaining parties are generally aware of their partner's gender, which may alter the strategic environment. We therefore alos look at how gender-revelation impacts posturing behavior.

In the treatment, subjects can condition bargaining behavior on their partner's gender. We find that this change in the strategic environment has a significant impact on bargaining behavior. The gender gap we observe in the control condition reverses in the treatment, where male subjects are more likely than female subjects to adopt a fair posture and less likely to adopt a strategic posture.

When we look at bargaining behavior round-by-round, we observe that the gender gap for both the strategic and fair posture is consistent over the rounds. However, while male subjects generally adopt a fair posture more often than female subjects in the treatment, a meaningful gender gap emerges only after subjects gain some experience of the behavior of male and female opponents. We see a similar pattern for the strategic posture, with a meaningful gender gap emerging after the first (approximately) 10 rounds.

Posturing behavior can affect the chances of reaching an agreement in the first stage, and Table 3 shows the frequency with which pairs reach agreement in stage 1, thereby avoiding the inefficiency of bargaining delays (the 290 pairs in which at least one of the partners was an induced type are excluded, leaving 910 pair observations).

Table 3: Agreements in stage 1

	Percentage of Agreements in Stage 1						
Pairs	**Male**		**Female**		**Mixed**		***Total***
	Mean	**Obs.**	**Mean**	**Obs.**	**Mean**	**Obs.**	
Control	18.095	105	28.788	66	32.642	193	*364*
Treatment	43.077	130	32.308	130	29.371	286	*546*
Total		*235*		*196*		*479*	*910*

From the first row, we see that male-male pairs have the lowest agreement rate in the control, where male subjects more often adopt the strategic posture and female subjects more often adopt the fair posture. On the other hand, the second row shows that in the treatment condition, where the gender gap is reversed, male-male pairs reach agreement more often than either mixed or female-female pairs.

Pairs not reaching agreement in stage 1 enter the second stage continuous-time concession game, where delays generate inefficiency. Using paired data on delays in stage 2, we observe that male-male pairs generate the greatest delays, but only in the treatment condition (where subjects are aware of their partner's gender) and only when at least one of the subjects in the pair adopted a strategic posture. This finding suggests that a strategic posture by a male subject does not provide a credible signal of commitment, especially against another male opponent. Male subjects who either anticipate these behavioral responses in stage 2 or learn about them from initial rounds, would have less incentive to adopt strategic posture in the treatment condition.

Finally, Table 4 shows how gender differences in bargaining behavior across the two stages affect total earnings of the subjects (in points). From the total 2400 observations, we exclude observations for induced types and for subjects when they are paired with an induced type, leaving 1820 observations.

Table 4: Points

Individuals	Average Points				
	Male		Female		Total
	Mean	**Obs.**	**Mean**	**Obs.**	
Control	13.416	403	12.742	325	728
Treatment	12.885	546	13.466	546	1092
Total		*949*		*871*	*1820*

The first row shows that, in the control, female subjects earn slightly less points than male subjects (approximately 12.7 points vs. 13.4 points). In the treatment condition, on the other hand, female subjects earn slightly more points (approximately 13.5 points vs. 12.9 points).

6 Regression analysis

In this section, we conduct regression analyses to complement our findings in Section 5, starting with individual demand data from stage 1 and then paired delay data from stage 2.

6.1 Posturing behavior

Table 5 provides the results of a regression analysis for the posturing behavior of subjects. The dependent variable in Column 1 is $demand^{15}$, which takes value 1 when a subject demands 15 and 0 otherwise. The dependent variable in Column 2 is $demand^{20}$, which takes value 1 when a subject demands 20 and 0 otherwise. We regress these dependent variables on the treatment dummy ($treatment_i$) and gender dummy ($male_i$; see Section 4.3), and the interaction of the treatment and gender dummy.[14] Since the dependent variables are binary, we estimate the following probit regression specification:

$$P\left(demand_{ir}^{15}=1|x_i\right) = \Phi\left(\alpha + \beta_t * treatment_i + \beta_m * male_i + \beta_{tm} * treatment_i * male_i\right), \quad (1)$$

where Φ is the cumulative distribution function of the standard normal distribution, and x_i is the vector of independent variables. For statistical inference, we calculate robust standard errors clustered at the session level, allowing for arbitrary correlation of errors across rounds and subjects in a session.[15] We exclude subjects who were in the role of induced type from the

[14] In Appendix A.1.1, we show that the results reported in Table 5 are robust to including controls for subject's age and major, as well as a dummy variable controlling for whether the session is gender-balanced or not. Since these controls have no significant effects, we omit them from the regression analysis to simplify the calculation and interpretation of marginal effects.

[15] Alternatively, we could also exploit the panel structure to estimate a panel probit regression with random effects. For comparison, we provide the results from a random effects specification in Appendix A.1.1. We opt

regression, leaving a sample size of 2100 observations (see Table 1 for details).

For both dependent variables—$demand^{15}$ in Column 1 and $demand^{20}$ in Column 2—Panel A reports the estimated coefficients. Panel B reports the marginal effects corresponding to the results discussed in Section 5.1 (Figure 3). Robust standard errors clustered at the session level are given in brackets. The p-values for the null hypothesis that the coefficient/marginal effect is equal to 0 are given in square brackets.

We find a significant gender gap in the control condition: relative to a male subject, a female subject is significantly more likely to demand 15 in the control, and significantly less likely to demand 20. On the other hand, in terms of the estimated marginal effects, female subjects are more likely to demand 20 in the treatment than male subjects, and less likely to demand 15. While the gender gaps for posturing behavior in the treatment are not statistically significant, the gender gaps in the treatment are significantly different from the gender gaps in the control. In particular, the p-value on "treatment effect for gender gap" indicates that we can reject the null hypothesis that the gender gap in the treatment is equal to the gender gap in the control. The regression results are therefore broadly in line with our analysis in Section 5: the gender-revelation treatment eliminates the significant gender gap observed in the control.

The estimated marginal treatment effects by gender are also in line with the qualitative findings in Section 5.1, but only the treatment effect for male subjects adopting a strategic posture is statistically significant: male subjects are significantly less likely to demand 20 in the treatment than in the control.

6.2 Bargaining delays

Using data for stage 2, we next look at how delays depend on the gender-composition of the pair. The baseline sample consists of all 627 pairs that entered stage 2, and in which neither partner was an induced type (see Table 2 for details).

Column 1 in Table 6 reports the results of a linear regression for the continuous dependent variable $Delay_{ij}$, which measures the time (in seconds) it took for one of the subjects i or j to concede after the pair ij entered the second stage:

$$\begin{aligned} Delay_{ij} = \alpha + \beta_T{*}Treatment_{ij} + \beta_M{*}Mixed_{ij} + \beta_F{*}Female_{ij} + \beta_{TM}{*}Treatment_{ij}{*}Mixed_{ij} \\ + \beta_{TF}{*}Treatment_{ij}{*}Female_{ij} + \beta_{d_i}{*}demand_i + \beta_{d_j}{*}demand_j + \varepsilon_{ij}, \end{aligned} \tag{2}$$

where $Treatment_{ij}$ is the treatment dummy; $Mixed_{ij}$ and $Female_{ij}$ are the gender-composition dummies described in Section 4.3; $demand_i$ (respectively, $demand_j$) is a continuous variable to

for the pooled probit regression with robust clustering at the session level as our main specification because it requires fewer assumption on the correlation of error terms and there is no ambiguity about the calculation of marginal effects.

Table 5: Demands in stage 1

	(1) $\mathbf{demand}_i^{15}$	(2) $\mathbf{demand}_i^{20}$
Panel A	Regression Coefficients	Regression Coefficients
constant (α)	0.0502	-0.567***
	(0.112)	(0.113)
	[0.653]	[0.000]
treatment (β_t)	-0.308	0.252
	(0.197)	(0.171)
	[0.119]	[0.142]
male (β_m)	-0.475***	0.613***
	(0.0546)	(0.164)
	[0.000]	[0.000]
treatment*male (β_{tm})	0.720***	-0.781***
	(0.263)	(0.259)
	[0.006]	[0.003]
Panel B	Marginal Effects	Marginal Effects
Gender gap in control	-0.185***	0.233***
$(\Phi(\alpha+\beta_m)-\Phi(\alpha))$	(0.021)	(0.062)
	[0.000]	[0.000]
Gender gap in treatment	0.097	-0.062
$(\Phi(\alpha+\beta_m+\beta_{tm})-\Phi(\alpha))$	(0.102)	(0.073)
	[0.342]	[0.398]
Treatment effect for gender gap	0.281***	-0.295***
$(\Phi(\alpha+\beta_m+\beta_{tm})$-$\Phi(\alpha+\beta_m))$	(0.104)	(0.096)
	[0.007]	[0.002]
Treatment effect for males	0.160	-0.204**
$(\Phi(\alpha+\beta_t+\beta_{tm})-\Phi(\alpha))$	(0.106)	(0.079)
	[0.132]	[0.010]
Treatment effect for females	-0.122	0.091
$(\Phi(\alpha+\beta_t)-\Phi(\alpha))$	(0.077)	(0.062)
	[0.114]	[0.144]
No. of Obs.	2100	2100

Note: Results from regression specification (1): probit regression of $demand^{15}$ (Column 1) or $demand^{20}$ (Column 2) on *treatment*, *male*, and $treatment * male$. Panel *A* provides regression coefficients; Panel *B* calculates marginal effects. Standard errors (in brackets) are clustered at the session level. P-values [in square brackets] are for the null-hypothesis that the coefficient/marginal effect is equal to 0. All numeric values are displayed up to 3 decimal places. Stars indicate significance: ** p < 0.050, *** p < 0.010.

control for the demand of subject i (respectively, j) in stage 1; and ε_{ij} is a random error term. The gender-composition dummy $Male_{ij}$ is the omitted variable. To assess how delays depend on strategic posturing in stage 1, we also conduct the analogous regression for two subsamples. Column 2 in Table 6 reports the results of an analogous regression specification but excluding those 100 observations where neither partner in the pair ij demanded 20 in stage 1 (leaving 523 observations), while Column 3 reports the results for the subsample excluding those 523 observations observations where at least one of the subjects in pair ij demanded 20 in stage 1 (leaving 100 observations).

For all three regressions (Column 1–3), Panel A reports the estimated regression coefficients, and Panel B calculates the gender-composition effects in the treatment, and also contains for comparison the relevant gender-composition effects in the control. Robust standard errors clustered at the session level are reported in brackets. The p-values for the null hypothesis that the coefficient/marginal effect is equal to 0 are given in square brackets.

From Column 1, the main finding is that male-male pairs experience significantly longer delays than mixed-gender or female-female pairs, but only in the treatment where the gender of the partners is revealed: relative to male-male pairs, the delays for mixed-gender and female-female pairs are significantly shorter in the treatment, but not statistically different in the control. The results are therefore in line with the qualitative findings we discuss in Section 5.2.

The results reported in Columns 2 and 3 suggest that the longer delays for male-male pairs in the treatment are related to the way that male subjects respond when another male subject adopts a strategic posture (mimicking an induced type by demanding 20 in stage 1).[16] In particular, when we focus on pairs where at least one partner demanded 20 in stage 1, Column 2 shows that the delays for male-male pairs in the treatment are even longer relative to mixed-gender and female-female pairs. However, when we focus on pairs where neither partner demanded 20 in stage 1, Column 3 indicates that there are no significant gender-composition effects.[17] The regression results for delays, where we include additional controls for demands in stage 1 and cluster errors at the session level, are therefore also broadly in line with our qualitative findings in Section 5.

[16] Similar to the findings in Column 1, there are no significant gender-composition effects in the control for the subsamples in Columns 2 and 3, which is anticipated as partner's gender is unknown in the control.

[17] There are substantially fewer observations in Column 3 because when neither partner demands 20 the pair often reaches agreement in stage 1 (e.g., when both demand 15). If anything, the point estimates suggest that male-male pairs have shorter delays than mixed-gender pairs in the treatment when neither partner initially demands 20, although the difference is not statistically significant.

Table 6: Delays in stage 2

	(1) Delay	(2) Delay with demand of 20	(3) Delay with demand not 20
Panel A	Regression Coefficients	Regression Coefficients	Regression Coefficients
Constant	-89.82***	-98.10**	-56.95
(α)	(22.89)	(36.24)	(45.36)
	[0.003]	[0.024]	[0.241]
Treatment	29.77**	35.39**	-13.64
(β_T)	(10.18)	(11.42)	(16.21)
	[0.017]	[0.013]	[0.422]
Mixed	1.975	6.189	-26.50*
(β_M)	(5.358)	(5.005)	(12.97)
	[0.721]	[0.248]	[0.071]
Female	-6.288	-6.439	4.896
(β_F)	(8.528)	(8.125)	(22.44)
	[0.480]	[0.448]	[0.832]
Treatment*Mixed	-28.88***	-45.93***	78.38
(β_{TM})	(7.513)	(11.83)	(41.02)
	[0.004]	[0.004]	[0.088]
Treatment*Female	-32.82**	-36.89**	-7.105
(β_{TF})	(11.62)	(12.82)	(24.83)
	[0.020]	[0.018]	[0.781]
demand$_i$	3.376***	2.315**	3.683*
(β_{d_i})	(0.996)	(0.737)	(2.006)
	[0.008]	[0.012]	[0.099]
demand$_j$	2.995**	4.517**	0.730
(β_{d_j})	(1.075)	(1.716)	(1.357)
	[0.021]	[0.027]	[0.604]
Panel B	Interaction Effects	Interaction Effects	Interaction Effects
Mixed pair effect in control	1.975	6.189	-26.495
(β_M)	5.358	5.005	12.966
	0.721	0.248	0.071
Female pair effect in control	-6.288	-6.439	4.896
(β_F)	8.528	8.125	22.435
	0.480	0.448	0.832
Mixed pair effect in treatment	-26.902***	-39.745***	51.889
($\beta_M + \beta_{TM}$)	(5.618)	(10.675)	(33.954)
	[0.001]	[0.005]	[0.161]
Female pair effect in treatment	-39.112***	-43.329***	-2.209
($\beta_F + \beta_{TF}$)	(9.467)	(10.583)	(14.627)
	[0.003]	[0.003]	[0.883]
No. of obs.	627	527	100

Note: Results from regression specification (2): linear regression of $Delay$ on $Treatment$, $Mixed$, $Female$, $Treatment * Mixed$, $Treatment * Female$, and control variables $demand_i$ and $demand_j$. Column 1 is for the sample of 627 pairs that reached stage 2 and where neither partner is an induced type; Column 2 is for the subsample of 527 pairs where at least one partner demanded 20 in stage 1; Column 3is for the subsample of 100 pairs where neither partner demanded 20 in stage 1. Panel A provides regression coefficients; Panel B calculates marginal effects. Standard errors (in brackets) are clustered at the session level. P-values [in square brackets] are for the null-hypothesis that the coefficient/marginal effect is equal to 0. All numeric values are displayed up to 3 decimal places. Stars indicate significance: ** p < 0.050, *** p < 0.010.

7 Conclusion

An essential feature of many bargaining problems is that parties can often try to bluff about their commitment to a specific bargaining outcome in order to induce concessions from their opponent. We conduct an experiment to investigate gender differences in the willingness and ability to bluff about one's commitment, and the impact of these strategic postures on bargaining delays.

Our design is based on Embrey et al.'s (2015) implementation of the bargaining with reputation model in Abreu and Gul (2000), where irrational types are induced in the lab to generate a bargaining problem with two-sided asymmetric information. Although the bargaining protocol is stylized, it provides a simple framework to uncover how willing subjects are to exploit information asymmetries by adopting a strategic posture. By implementing the bargaining experiment with two conditions, which differ in whether gender is revealed (treatment) or not (control), we obtain new insights into the way that men and women try to exploit information asymmetries in a dynamic bargaining problem. In particular, we are able to address two main research questions. First, when gender is not revealed, we find substantial gender differences in posturing behavior: men are much more likely to adopt a strategic posture, while women are much more likely to adopt a fair posture. Second, we find that revealing gender seems to alter the strategic environment in a meaningful way. In particular, revealing gender mitigates (or even reverses) the gender gap in posturing behavior: if anything, women are more likely than men to adopt a strategic posture and less likely to adopt a fair one.

Looking at behavior in the continuation game when no agreement is reached in stage 1, we also observe that male-male pairs experience significantly longer bargaining delays than mixed-gender or female-female pairs, but only in the treatment condition where gender is revealed and only in those pairs where one of the subjects mimics the irrational type. This finding suggests that strategic postures do not provide a credible signal of commitment for male subjects, especially against a male opponent.

References

Abreu, D. and F. Gul (2000): "Bargaining and Reputation," *Econometrica*, 68, 85–117.

Aguiar, F., P. Brañas-Garza, R. Cobo-Reyes, N. Jimenez, and L. M. Miller (2009): "Are women expected to be more generous?" *Experimental Economics*, 12, 93–98.

Andreoni, J. and L. Vesterlund (2001): "Which is the Fair Sex? Gender Differences in Altruism," *The Quarterly Journal of Economics*, 116, 293–312.

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

Ayres, I. and P. Siegelman (1995): "Race and Gender Discrimination in Bargaining for a New Car," *American Economic Review*, 85, 304–321.

Azmat, G. and B. Petrongolo (2014): "Gender and the Labor Market: What Have We Learned from Field and Lab Experiments?" *Labour Economics*, 30, 32–40.

Babcock, L. and S. Laschever (2009): *Women don't ask: Negotiation and the gender divide*, Princeton University Press.

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

Bock, O., I. Baetge, and A. Nicklisch (2014): "hroot – Hamburg registration and organization online tool," *European Economic Review*, 71, 117–120.

Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2019): "Beliefs about Gender," *American Economic Review*, 109, 739–773.

Bowles, H. R. (2014): "Psychological Perspectives on Gender in Negotiation," in *The SAGE Handbook of Gender and Psychology*, ed. by M. K. Ryan and N. R. Branscombe, SAGE Publications Ltd, chap. 28.

Byrnes, J. P., D. C. Miller, and W. D. Schafer (1999): "Gender differences in risk taking: a meta-analysis." *Psychological bulletin*, 125, 367.

Card, D., A. R. Cardoso, and P. Kline (2016): "Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women," *The Quarterly Journal of Economics*, 131, 633–686.

Castillo, M., R. Petrie, and M. T. an Lise Vesterlund (2013): "Gender Differences in Bargaining Outcomes: A Field Experiment on Discrimination," *Journal of Public Economics*, 99, 35–48.

Coffman, K. B. (2014): "Evidence on Self-Stereotyping and the Contribution of Ideas," *The Quarterly Journal of Economics*, 129, 1625–1660.

Croson, R. and U. Gneezy (2009): "Gender Differences in Preferences," *Journal of Economic Literature*, 47, 1–27.

Damiano, E., H. Li, and W. Suen (2012): "Optimal Deadlines for Agreements," *Theoretical Economics*, 7, 257–393.

Dittrich, M., A. Knabe, and K. Leipold (2014): "Gender Differences in Experimental Wage Negotiations," *Economic Inquiry*, 52, 862–873.

Dreber, A. and M. Johannesson (2008): "Gender Differences in Deception," *Economics Letters*, 99, 197–199.

Eckel, C., A. C. M. de Oliveira, and P. J. Grossman (2008): "Gender and Negotiation in the Small: Are Women (Perceived to Be) More Cooperative than Men?" *Negotiation Journal*, 24, 429–445.

Eckel, C. C. and P. J. Grossman (1998): "Are Women Less Selfish than Men?: Evidence from Dictator Experiments," *The Economic Journal*, 108, 726–735.

——— (2001): "Chivalry and Solidarity in Ultimatum Games," *Economic Inquiry*, 39, 171–188.

Embrey, M., G. R. Frechette, and S. F. Lehrer (2015): "Bargaining and Reputation: An Experiment on Bargaining in the Presence of Behavioural Types," *Review of Economic Studies*, 82, 608–631.

Fang, H. and A. Moro (2011): "Theories of Statistical Discrimination and Affirmative Action: A Survey," in *Handbook of Social Economics*, Elsevier, vol. 1, 133–200.

Fanning, J. (2016): "Reputational Bargaining and Deadlines," *Econometrica*, 84, 1131–1176.

——— (2018): "No compromise: Uncertain costs in reputational bargaining," *Journal of Economic Theory*, 175, 518–555.

Farjam, M. (2019): "On whom would I want to depend; humans or computers?" *Journal of Economic Psychology*, 72, 219–228.

Fershtman, C. and U. Gneezy (2001): "Strategic Delegation: An Experiment," *RAND Journal of Economics*, 32, 352–368.

Fershtman, C. and E. Kalai (1997): "Unobserved Delegation," *International Economic Review*, 38, 763–774.

Fischbacher, U. (2007): "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics*, 10, 171–178.

Genesove, D. and C. J. Mayer (1997): "Equity and Time to Sale in the Real Estate Market," *American Economic Review*, 87, 255–269.

Goldberg, P. K. (1996): "Dealer Price Discrimination in New Car Purchases: Evidence from the Consumer Expenditure Survey," *Journal of Political Economy*, 104, 622–654.

Harding, J., S. Rosenthal, and C. F. Sirmans (2003): "Estimating Bargaining Power in the Market for Existing Homes," *The Review of Economics and Statistics*, 85, 178–188.

Harless, D. W. and G. E. Hoffer (2002): "Do Women Pay More for New Vehicles? Evidence from Transaction Price Data," *American Economic Review*, 92, 270–279.

Hernandez-Arenaz, I. and N. Iriberri (2018): "Women ask for less (only from men): Evidence from bargaining in the field," *Journal of Economic Behavior and Organization*, 152, 192–214.

Kray, L. and L. Thompson (2004): "Gender stereotypes and negotiation performance: An examination of theory and research," *Research in organizational behavior*, 26, 103–182.

Kwiek, M. (2011): "Reputation and Cooperation in the Repeated Second-Price Auctions," *Journal of the European Economic Association*, 9, 982–1001.

Myerson, R. B. (1991): *Game Theory: Analysis of Conflict*, Harvard University Press.

Niederle, M. (2014): "Gender," *NBER Working Paper 20788.*

Niederle, M. and L. Vesterlund (2007): "Do Women Shy Away from Competition? Do Men Compete Too Much?" *Quarterly Journal of Economics*, 122, 1067–1101.

Özyurt, S. (2015): "Searching for a Bargain: Power of Strategic Commitment," *American Economic Journal: Microeconomics*, 7, 320–353.

Pondorfer, A., T. Barsbai, and U. Schmidt (2016): "Gender differences in stereotypes of risk preferences: Experimental evidence from a matrilineal and a patrilineal society," *Management Science*, 63, 3268–3284.

Povel, P. (2005): "Multiple Banking as a Commitment Not to rescue," *Research in Finance*, 21, 175–199.

Roth, A. E. and M. W. K. Malouf (1979): "Game-Theoretic Models and the Role of Information in Bargaining," *Psychological Review*, 86, 574–594.

Roth, A. E., V. Prasnikar, M. Okuno-Fujiwara, and S. Zamir (1991): "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study." *The American Economic Review*, 81.

Rubinstein, A. (1982): "Perfect Equilibrium in a Bargaining Model," *Econometrica*, 50, 97–109.

Schotter, A., W. Zheng, and B. Snyder (2000): "Bargaining Through Agents: An Experimental Study of Delegation and Commitment," *Games and Economic Behavior*, 30, 248–292.

Seagraves, P. and P. Gallimore (2013): "The Gender Gap in Real Estate Sales: Negotiation Skill or Agent Selection?" *Real Estate Economics*, 41, 600–631.

Solnick, S. J. (2001): "Gender Differences in the Ultimatum Game," *Economic Inquiry*, 39, 189–200.

Sutter, M., R. Bosman, M. G. Kocher, and F. van Winden (2009): "Gender pairing and bargaining – Beware the same sex!" *Experimental Economics*, 12, 318–331.

A Appendix

Appendix A.1 provides additional data analysis, and Appendix A.2 provides our experimental instructions, translated from German, along with screenshots.

A.1 Additional data analysis

To complement the findings discussed in Section 5 and 6, Appendix A.1.1 presents the results of a number of robustness regressions; Appendix A.1.2 shows that there is no observable influence of partner's gender on posturing behavior; and Appendix A.1.3 considers two alternative approaches to account for demands other than 15 or 20 in the analysis of posturing behavior. Finally, in Appendix A.1.4, we illustrate average delays in stage 2 when we divide the sample of pairs into three categories—rather than two—in terms of the demands in stage 1.

A.1.1 Robustness

Table A1 presents a robustness analysis for our regression results on posturing behavior in Section 6. For comparison, Columns 1 and 2 provide the regression results from the probit regression specification (1) reported in Table 5.

Columns 3 and 4 present the regression results when we add additional controls for subject's age, a dummy variable taking value 1 if the subject is an economics or business major, and a dummy variable taking value 1 if the session is gender-balanced. All other details of the regression specification are as in the baseline (pooled probit with robust standard errors clustered

Table A1: Robustness checks for demands in stage 1

	Baseline		Baseline w. additional controls		Panel probit w. random effects	
	(1)	(2)	(3)	(4)	(5)	(6)
	demand$_i^{15}$	**demand**$_i^{20}$	**demand**$_i^{15}$	**demand**$_i^{20}$	**demand**$_i^{15}$	**demand**$_i^{20}$
constant	0.0502	-0.567***	0.207	-0.720	0.0577	-0.715***
(α)	(0.112)	(0.113)	(0.611)	(0.545)	(0.172)	(0.176)
	[0.653]	[0.000]	[0.735]	[0.187]	[0.737]	[0.000]
treatment	-0.308	0.252	-0.426	0.247	-0.456	0.207
(β_t)	(0.197)	(0.171)	(0.264)	(0.189)	(0.295)	(0.258)
	[0.119]	[0.142]	[0.107]	[0.193]	[0.122]	[0.424]
male	-0.475***	0.613***	-0.496***	0.601***	-0.616***	0.778***
(β_m)	(0.0546)	(0.164)	(0.0721)	(0.148)	(0.0534)	(0.214)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
treatment*male	0.720***	-0.781***	0.722***	-0.777***	0.978**	-1.061***
(β_{tm})	(0.263)	(0.259)	(0.276)	(0.262)	(0.403)	(0.368)
	[0.006]	[0.003]	[0.009]	[0.003]	[0.015]	[0.004]
age			0.00148	0.00691		
(β_{age})			(0.0268)	(0.0222)		
			[0.956]	[0.756]		
economics			-0.0345	-0.00609		
(β_{econ})			(0.0908)	(0.143)		
			[0.704]	[0.966]		
gender-balance			-0.349	-0.00397		
(β_{bal})			(0.257)	(0.137)		
			[0.175]	[0.977]		
No. of Obs.	2100	2100	2100	2100	2100	2100

Note: Results from regression specification (1): probit regression of $demand^{15}$ (Column 1) or $demand^{20}$ (Column 2) on *treatment*, *male*, and $treatment * male$; the same specification regression of $demand^{15}$ (Column 3) or $demand^{20}$ (Column 4) but with additional controls (*age* of the subject, *economics* dummy taking a value of 1 if subject is an economics major, $gender - balance$ dummy taking a value of 1 if the session is gender-balanced); the panel probit regression (with random effects) of $demand^{15}$ (Column 5) or $demand^{20}$ (Column 6). Standard errors (in brackets) are clustered at the session level. P-values [in square brackets] are for the null-hypothesis that the coefficient/marginal effect is equal to 0. All numeric values are displayed up to 3 decimal places. Stars indicate significance: ** $p < 0.050$, *** $p < 0.010$.

at the session level). We see that all additional controls are not significant, and including the controls has almost no impact on the regression coefficients or standard errors relative to the baseline regression.

Columns 5 and 6 present the regression results from a panel probit regression with random effects, and robust standard errors clustered at the session level. For direct comparison with the baseline, we omit the additional controls, which are also insignificant in the random effects specification. Again, we see that the regression coefficients, standard errors, and significance are very similar to the baseline regression.

A.1.2 Partner gender effect

In the treatment condition, subjects can condition their choice of initial posture on the gender of their partner. To assess whether partner's gender impacts posturing behavior, we focus on the treatment condition where we have 1260 individual demand observations (which excludes

Table A2: Partner gender effects for demands in stage 1

	(1) $\textbf{demand}_i^{15}$	(2) $\textbf{demand}_i^{20}$
Panel A	Regression Coefficients	Regression Coefficients
constant	-0.214	-0.353***
(α)	(0.149)	(0.133)
	[0.151]	[0.008]
$\boldsymbol{male}_i$	0.180	-0.100
(β_{m_i})	(0.267)	(0.221)
	[0.501]	[0.650]
$\boldsymbol{male}_j$	-0.0837	0.0710**
(β_{m_j})	(0.0683)	(0.0314)
	[0.221]	[0.024]
$\boldsymbol{male}_i * \boldsymbol{male}_j$	0.130***	-0.136***
($\beta_{m_i m_j}$)	(0.0411)	(0.0413)
	[0.001]	[0.001]
Panel B	Marginal Effects	Marginal Effects
Gender gap with female partner	0.071	-0.037
$(\Phi(\alpha + \beta_{m_i}) - \Phi(\alpha))$	(0.106)	(0.081)
	[0.502]	[0.649]
Gender gap with male partner	0.122	-0.087
$(\Phi(\alpha + \beta_{m_i} + \beta_{m_i m_j}) - \Phi(\alpha))$	(0.104)	(0.069)
	[0.243]	[0.208]
Partner gender effect for males	0.019	-0.023
$(\Phi(\alpha + \beta_{m_j} + \beta_{m_i m_j}) - \Phi(\alpha))$	(0.022)	(0.016)
	[0.400]	[0.165]
Partner gender effect for females	-0.032	0.027**
$(\Phi(\alpha + \beta_{m_j}) - \Phi(\alpha))$	(0.026)	(0.012)
	[0.206]	[0.026]
No. of Obs.	1260	1260

Note: Results from regression specification A.1: probit regression of $demand^{15}$ (Column 1) or $demand^{20}$ (Column 2) on $male_i$, $male_j$, and $male_i * male_j$. Standard errors (in brackets) are clustered at the session level. P-values [in square brackets] are for the null-hypothesis that the coefficient is equal to 0. All numeric values are displayed up to 3 decimal places. Stars indicate significance: ** $p < 0.050$, *** $p < 0.010$.

induced types). Table A2 reports results from the following probit specification:

$$P\left(demand_{ir}^{z}{=}1|x_i\right) = \Phi\left(\alpha + \beta_{m_i} * male_i + \beta_{m_j} * male_j + \beta_{m_i m_j} * male_i * male_j\right), \qquad \text{(A.1)}$$

where $z \in \{15, 20\}$, and the regressors are a gender dummy for the subject $male_i$, a gender dummy for the subjects partner $male_j$ (= 1 if subject i's partner in round r is male and 0 otherwise), as well as an interaction term. Panel A reports the coefficients and Panel B the marginal effects, together with robust standard errors clustered at the session level and p-values for a test of the null hypothesis that the coefficient/marginal effect is equal to 0.

Column 1 reports the results for the fair posture ($demand^{15}$) and Column 2 reports the results for the strategic posture ($demand^{20}$). The first two rows of Panel B give marginal effects for the gender gap against a female (respectively, male partner), i.e., whether male subjects choose 15 (Column 1) or 20 (Column 2) more often than female subjects when matched with a female (first row) or male (second row) partner. The third and fourth rows of Panel B show marginal effects for partner gender effects, i.e., whether a male (third row) or female (fourth row) subject choose 15 (Column 1) or 20 (Column 2) more often when matched with a male partner than if they are matched with a female partner.

The results show that the effects are all either statistically or economically insignificant. The only statistically significant effect is the partner gender effect for females with $demand^{20}$, which indicates that female subjects are more likely to adopt a strategic posture against a male opponent than a female opponent. However, the effect size is an order of magnitude smaller than the effects we discuss in Sections 5.1 and 6.1. From the small effect sizes and lack of significance, we infer that partner's gender does not appear to be an important factor in explaining posturing behavior in our experiment.

A.1.3 Multinomial probit

For our analysis of posturing behavior, we consider separate regressions for the binary variables $demand^{15}$ and $demand^{20}$ because both have a direct economic interpretation: demanding 20 in stage 1 is mimicking the behavior of an induced type and corresponds to the equilibrium prediction if induced types are the only irrational type; demanding 15 can be interpreted as reflecting a natural fairness norm, as the bargaining protocol is symmetric. More than 80% of demand observations in both the control and treatment correspond to either the fair or strategic posture (see Section 4.2). However, there are subjects whose initial demand is not 15 or 20 in at least some rounds, and we remain largely agnostic about the rationale for departing from the fair or strategic posture.

In Table A3, we present two alternative approaches to explicitly account for demands other than 15 or 20. For comparison, Columns 1 and 2 present our baseline regression results from

Section 6.1. Column 3 presents results when we simply exclude the 403 demand observations not equal to 15 or 20, and run our baseline probit regression for the binary variable $demand^{15}$. With the exclusion of demands other than 15 or 20, $demand^{15}$ and $demand^{20}$ are perfectly negatively correlated and one can calculate marginal effects for both the fair and strategic posture from the coefficients, which are very similar to the coefficients for $demand^{15}$ in the baseline.

Column 4 and 5 are from a single multinomial regression with a categorical dependent variable that takes value 0 when demand is 15, value 1 when demand is 20, and value 3 when demand is other than 15 or 20. Demand 15 is therefore the baseline category. In the multinomial specification, marginal effects are always relative to the baseline category, and it is therefore difficult to isolate gender and treatment effects for the strategic (or fair) posture in isolation. However, we see from Column 4 that gender gap for demand of 20, relative to demand of 15, is significant in the control and reverses in the treatment, broadly in line with our findings for the separate probit regressions in Section 6.1.

Table A3: Demand other in stage 1

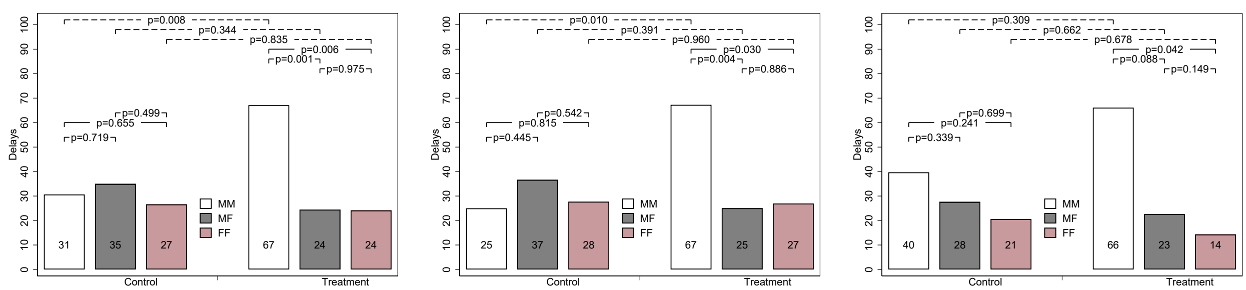
	Baseline		**Exclude other demands**	**Multinomial probit w. base: demand$_i^{15}$**	
	(1) **demand$_i^{15}$**	(2) **demand$_i^{20}$**	(3) **demand$_i^{15}=1$**	(4) **demand$_i^{20}$**	(5) **demand$_i^{Other}$**
constant (α)	0.0502	-0.567***	0.374***	0.492***	-0.773***
	(0.112)	(0.113)	(0.130)	(0.171)	(0.147)
	[0.653]	[0.000]	[0.004]	[0.004]	[0.000]
treatment (β_t)	-0.308	0.252	-0.338*	0.445*	0.333
	(0.197)	(0.171)	(0.205)	(0.268)	(0.317)
	[0.119]	[0.142]	[0.099]	[0.097]	[0.293]
male (β_m)	-0.475***	0.613***	-0.645***	0.856***	0.169
	(0.0546)	(0.164)	(0.114)	(0.151)	(0.247)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.493]
treatment*male (β_{tm})	0.720***	-0.781***	0.893***	-1.183***	-0.475
	(0.263)	(0.259)	(0.285)	(0.380)	(0.419)
	[0.006]	[0.003]	[0.002]	[0.002]	[0.258]
No. of Obs.	2100	2100	1697	2100	

Note: Results from regression specification (1): probit regression of $demand^{15}$ (Column 1) or $demand^{20}$ (Column 2) on *treatment*, *male*, $treatment*male$, *age* and *economics*. Excluding the observations where the subject demanded other than $demand^{15}$ or $demand^{20}$, a single demand variable is constructed with $demand^{15}=1$ and 0 when $demand^{20}=1$ (Column 3). Results from multinomial regression with the dependent variable of demand as a categorical variable with 3 categories: Baseline category of 0 is $demand^{15}$, category 2 is $demand^{20}$ and category 3 is $demand^{Others}$ (Column 4-5). Standard errors (in brackets) are clustered at the session level. P-values [in square brackets] are for the null-hypothesis that the coefficient is equal to 0. All numeric values are displayed up to 3 decimal places. Stars indicate significance: ** p < 0.050, *** p < 0.010.

A.1.4 Delays conditional on stage 1 demands

In Sections 5.2 and 6.2, we analyze delays for the subsample of 100 pairs where neither partner adopts a strategic posture in stage 1 and the subsample of 527 pairs where at least one partner demands 20 in stage 1. The second subsample could be divided further into (i) the 407 pairs where exactly one partner demands 20 and (ii) the 120 pairs where both subjects demand 20. However, Figure A1 illustrates that there are essentially no difference in average delays, broken down by condition (control or treatment) and gender-composition of the pair, between the subsamples (i) and (ii). For comparison, Figure 1(a) illustrates the average delays in the control and treatment for each of the gender-pairings when at least one partner demands 20 (Figure 7(b) in Section 5.2). Figure 1(b) illustrates the conditional average delays when exactly one partner demands 20, and Figure 1(c) illustrates the conditional average delays when both partners demand 20. We therefore opt to merge these categories to improve the power of the statistical inferences since, in particular, there are relatively few observations (only 8) for male-male pairs both demanding 20 in the treatment.

Figure A1: Delays in stage 2 conditional on posturing in stage 1



(a) At least one demand of 20 (b) Exactly one demand of 20 (c) Both demand 20

Note: Subfigure (a) displays the average delays in stage 2 by gender composition of the pairs when at least one subject demanded 20 in stage 1 (same as subfigure 7(b)). Subfigure (b) displays the average delays in stage 2 by gender composition of the pairs when exactly one subject demanded 20 in stage 1. Subfigure (c) displays the average delays in stage 2 by gender composition of the pairs when both subjects demanded 20 in stage 1. Conditional means are based on the observations summarized in Panel B of Table 2, rounded to the nearest integer. Top horizontal bars show p-values for a t-test of the null hypothesis that the conditional means are equal.

A.2 Experimental screenshots and instructions

Figure A.2.1: Screenshot (initial demand)

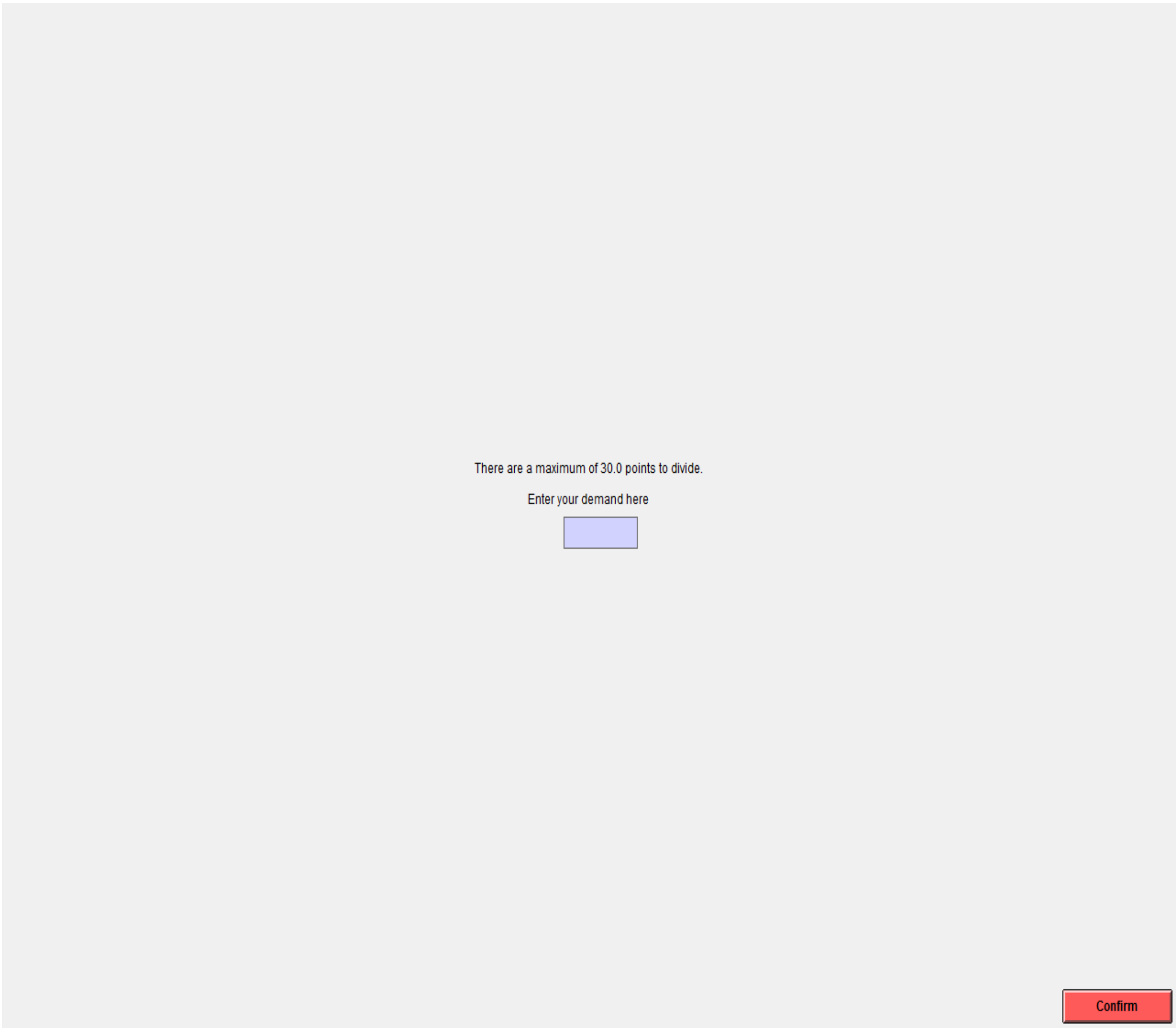
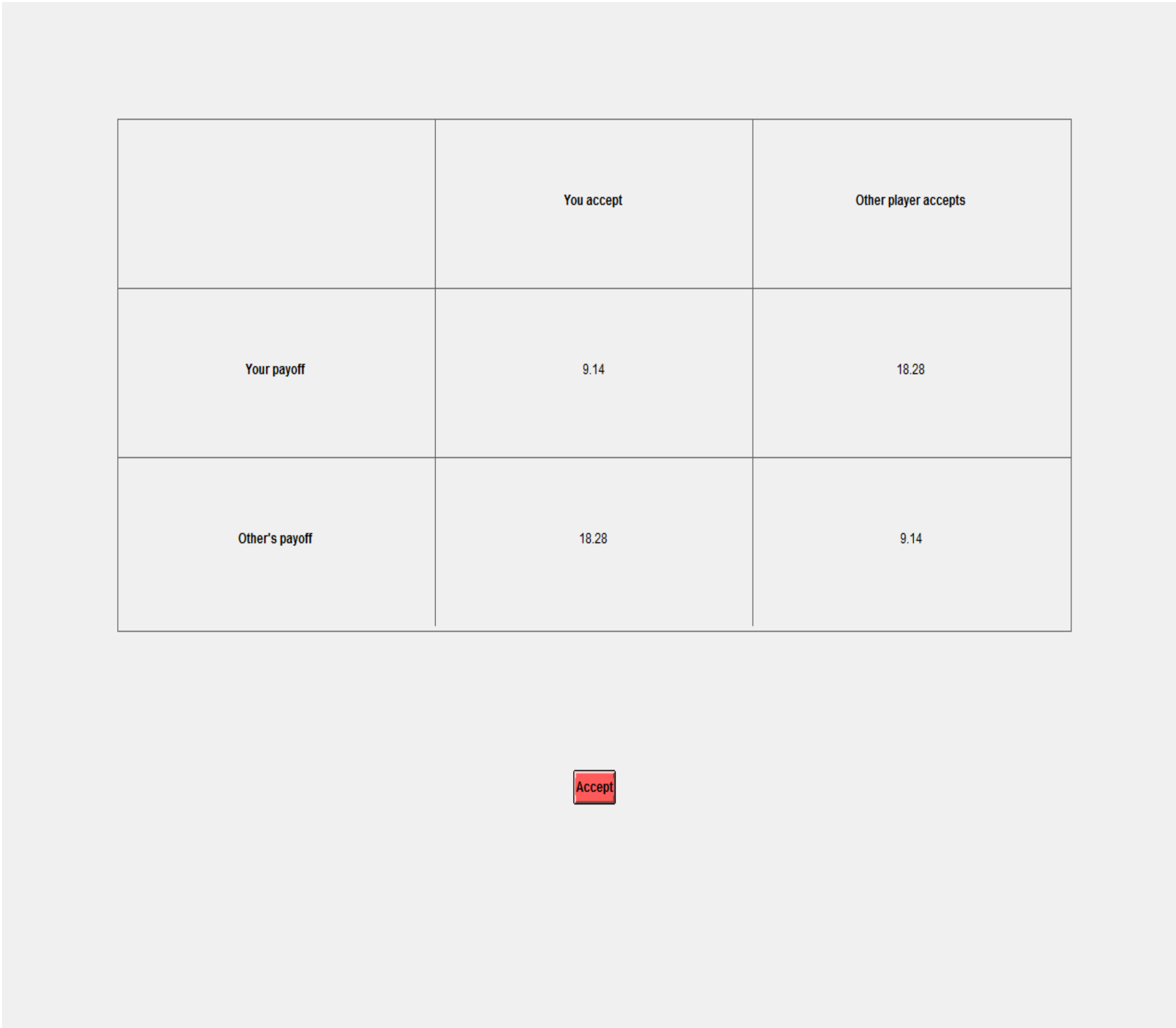


Figure A.2.2: Screenshot (stage 2)



	You accept	Other player accepts
Your payoff	9.14	18.28
Other's payoff	18.28	9.14

Instructions Part 1 (Control)

In front of you, there is an envelope. In this envelope there a pseudonym. Every player receives their own pseudonym (e.g., **“player Berlin”**) and keeps their pseudonym throughout the entire experiment.

In part 2 you will play against other subjects. Players will receive information about the pseudonym of the other players with whom they are paired. This is why you need to type in your pseudonym in part 1.

As soon as any questions have been answered, part 1 will begin on your computers. On the screen, you will a space for typing. Type in your pseudonym (e.g., **“player Berlin”**) and press on **“Continue”**.
As soon as all players have typed in their pseudonym you will be directed to a new screen in which you are asked to type a password. Please then open your cabin door, you will receive the instructions to part 2.

In Summary:

1. Type in your pseudonym. For example **“player Berlin”**.
2. Press on **“Continue”**.
3. Open your cabin door as soon as a password is to be entered.

Are there any questions?

Instructions Part 1 (Treatment)

In front of you, there is an envelope. In this envelope there a pseudonym. Every player receives their own pseudonym (e.g., **"player Berlin"**) and keeps their pseudonym throughout the entire experiment.

In part 2 you will play against other subjects. Players will receive information about the pseudonym of the other players with whom they are paired. This is why you need to record the audio file in part 1.

To do so, put on your headset and make sure the microphone is in front of your mouth. As soon as any questions have been answered, part 1 will begin on your computers. On the screen, you will see a button "**begin recording**". After you have pressed the button wait until you see the message "**no cam**" and then say your pseudonym (e.g., **"player Berlin"** clearly into the microphone. Afterwards stay quiet and say nothing further, the recording will end automatically As soon as all players have recorded their pseudonym you will be directed to a new screen in which you are asked to type a password. Please then open your cabin door, you will receive the instructions to part 2.

In Summary:

1. Put on your headsets and place the microphone in front of your mouth.
2. Press the button "**begin recording"** and stay quiet.
3. Wait until the message "**no cam**" in displayed (approx. 2-3 seconds).
4. Say clearly your pseudonym. For example **"player Berlin"**.
5. Stay quiet and leave your headset on.
6. Open your cabin door as soon as a password is to be entered.

Are there any questions?

Instructions- Part 2

There are a total of 16 players in this experiment, you and 15 others. There are two types, Diamond and Spade. Each of the 16 players will learn their type at the start of the experiment and everyone keeps their type throughout the entire experiment. There are 14 type Diamond and 2 are type Spade. Which type you are is determined at random.

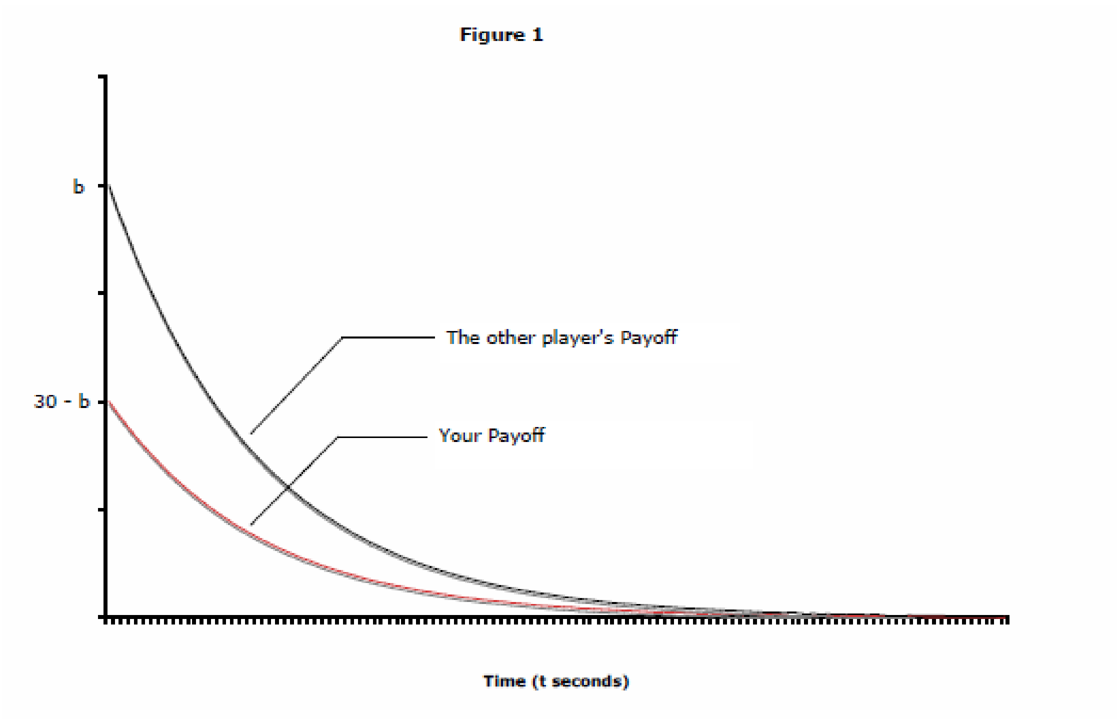
As a player of type Diamond, you will make decisions over 15 periods. At the beginning of each period, you will be matched with a randomly assigned player. That player will be either another player of type Diamond or one of type Spade (more on a type Spade later). At the start of each period, you will hear the pseudonym of the other player. For this reason, you should keep your headsets on throughout the experiment. **[This instruction differed in the control condition where the instruction was: At the start of each period, you will see the pseudonym of the other player.]** During each period, your task is to divide 30 points between yourself and the other player you are matched with.

Each period has up to two stages:

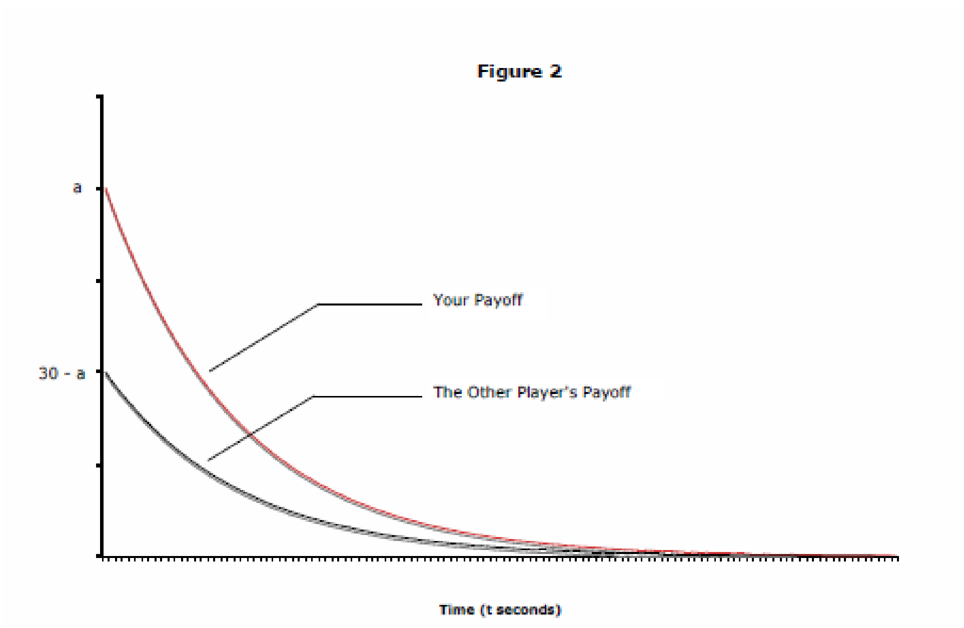
Stage 1: You place an announcement for the number of points that you want for yourself out of the 30 (denote this by a). Simultaneously, the other player will make an announcement for the number of points they want for themselves (denote this by b).

- If the two announcements sum to 30 or less, then you will receive your announcement plus half of what is left over (30 minus the sum of the two announcements) and the period will end. In other words, you will receive a + (30-a-b)/2 points and the other player receives b + (30-a-b)/2.
- If the two announcements sum to more than 30, then you move on to the second stage.

Stage 2: You can now either accept the other player's announcement or wait until they accept your announcement. Accepting their announcement immediately means that you receive 30 – b points for that period. However, the longer you wait the less your points are worth. Approximately, points decrease at a rate of 1% per second. More precisely, if you accept the other player's announcement after t seconds, you will receive $(30 - b)$ x $(0.99)^t$ and the other player will receive b x $(0.99)^t$. The following graph illustrates this:



If on the other hand, the other player accepts your offer after t seconds, you will receive a x $(0.99)^t$ and the other player will receive $(30 - \text{a})$ x $(0.99)^t$. The following graph illustrates this:



Your computer screen will display the points you and the other player would receive if you were to accept, or if they were to accept your announcement at different points in time. Once either you or the other player has accepted, or the value of the points have reached zero, the period is over.

A few examples might help your understanding. These are not meant to be realistic:

1. In the first stage, you announce 1.5 and the other player announces 3.5. Since 1.5 + 3.5 = 5, which is smaller than 30, the period ends and you receive 1.5 + (30 - 5)/2 = 14 points. If instead the other player had announced 23.5, then you would have received 1.5 + (30 - 25)/2 = 4 points.

2. In the first stage, you announce 15 and the other player announces 23. Since 15 + 23 = 38, which is greater than 30, you go to the second stage. In the second stage, the other accepts your announcement after 1 second. You get 15 x $(0.99)^1$ = 14.85 points. If instead, the other player does not accept immediately and you accept after 10 seconds, then you obtain $(30 - 23)$ x $(0.99)^{10}$ = 6.33 points.

3. In the first stage, you announce 25 and the other player announces 5. Since 25 + 5 = 30, the period ends and you obtain 25 points.
As you can see there are many possibilities.

When every pair has finished this task, the next period begins. You will be matched with a randomly assigned player in the next period. The task in the next period is exactly the same as the one just described (apart from that you will be playing with a new player).
The experiment consists of 15 such periods.

Players of type Spade do the same thing every period. Their strategy is as follows. In the first stage, the Spade player will always announce that they want 20 points. If the period goes to the second stage (that is the announcements are incompatible), the Spade player will never accept the offer of the other player. At the beginning of each period, The Diamond player has a 2/15 chance of being matched to a Spade player.

Once the 15 periods have been completed, the total number of points you have earned will be displayed (denote this by P). These points determines the odds of winning a prize in your lottery. Your lottery has the following structure:

- The odds of winning are given by the number of points you earned throughout the experiment divided by the total number of points available. Since there are 15 periods and there are 30 points available in each period, the total number of points available is given by 15 x 30 = 450. Thus the odds of winning are P/450.
- The prize is 20 euro.
- That is, you have P/450 chance of winning the prize and 1 - P/450 chance of receiving 0.

In summary, your earning from this session is comprised of a 10 euro participation fee and the outcome of your lottery. The probabilities associated with your lottery depend on the number of points you have earned throughout the session. You can earn either 0 or 20 from the lottery.

Are there any questions?

Summary

Before we start, let me remind you that:

- After a period is finished, you will be matched to a randomly assigned new player for the next period. You will hear the pseudonym of your partner via your headset.
- In each period, you and another player will make announcements to divide 30 points between both of you. If the sum of your two announcements is less than 30 the period ends. If the sum of the two announcements is 30 or more you move to a second stage. In the second stage, the points decrease in value until either you or the other player accepts the announcement made by the other party, at which point the period ends.
- At the end of the session, your earnings are determined by a lottery with probabilities that depend on the number of points you have earned throughout the experiment. You can earn either 0 or 20 from the lottery. In addition you will receive a 10 euro show-up fee.

Good Luck.