Commitment and Threats in Adversarial Bargaining: Insights from a Plea Bargaining Experiment

Pablo Cuellar*, Lucas Rentschler†

May 2023

Abstract

We study the effect of commitment on threats in adversarial bargaining—situations in which a prosecutor demands payment from a defendant under the threat of conflict if the demand is rejected. We focus on plea bargaining, where a prosecutor files charges against a defendant before a round of plea bargaining. The charges affect payoffs if an agreement is not reached, and the case proceeds to trial. The expected benefit for the prosecutor at trial is maximized at a lower charges level than those that maximize the responder's loss. Just before the trial starts, the prosecutor can reduce the charges up to a certain point depending on their level of commitment power. The experimental data show that participants with sufficient commitment power used it to gain bargaining power by overcharging the defendant. However, they do so more conservatively than the theory predicts.

Keywords: C78, C92, D74
JEL Codes: Adversarial bargaining, credible threats, conflict, experiments

*Utah State University. Email: pablo.cuellar@usu.edu
†Utah State University. Email: lucas.rentschler@usu.edu

1 Introduction

A credible commitment to a threat plays a vital role in increasing bargaining power in negotiations, often achieved through public actions that are difficult to retract. In the context of plea bargaining—a specific form of adversarial bargaining involving demands for payment under the threat of initiating detrimental actions—prosecutors offer plea deals to defendants, stipulating potential punishments. While the prosecutor prefers the harshest possible punishment, the defendant seeks the most lenient one. To bolster bargaining power, prosecutors may choose to overcharge defendants, filing multiple charges that coerce them into accepting severe plea deals to avoid the risks and consequences of a trial.

However, the credibility of the threat posed by high charges depends on whether the prosecutor will maintain them once the trial commences. If both the prosecutor and defendant are aware that proving the charges in court will be challenging, the prosecutor faces potential reputational costs if the case is lost at trial. Consequently, the prosecutor might prefer to lessen the severity of charges before the trial begins if a plea deal is rejected. If this is anticipated by the defendant, the prosecutor losses bargaining power.

In this paper, we present a one-period adversarial bargaining model that investigates the use of an exogenous imperfect pre-commitment device, such as the reputational cost associated with scaling down charges, to establish credibility in the threat posed by prosecutors. By threat, we refer to the possible punishment given by the charges files before the plea bargaining starts. Our theoretical prediction suggests that prosecutors employ this pre-commitment power to credibly commit to a high threat that maximizes the defendant's loss. Notably, such a threat is risky, as the prosecutor incurs significant losses if an agreement is not reached.

To test our predictions, we conduct a laboratory experiment that explores whether prosecutors utilize their commitment power to pose credible threats by choosing high charges, thereby placing defendants in risky positions. Furthermore, we examine whether this commitment power translates into higher payoffs compared to situations without commitment power.

The experimental results qualitatively confirm the predictions of our theory. Prosecutors commit to relatively high threats by choosing high charges and exposing themselves to the risk of going to trial and losing. However, the actual charges chosen are not as high as the theory predicts. Additionally, the results indicate that while prosecutors can successfully commit to a credible threat, it does not necessarily lead to higher payoffs.

To analyze this dynamic, we present an adversarial bargaining model consisting of two stages: plea bargaining and trial. In the plea bargaining stage, prosecutors choose charges

and subsequently offer a plea deal to the defendant. If the defendant accepts, the game concludes. However, if the plea deal is rejected, the trial stage ensues. The trial is a lottery that assigns payoffs depending on whether the defendant is found guilty or not guilty, which depends on the charges. Before the payoffs materialize, the prosecutor retains the option to reduce charges up to a predetermined exogenous point, representing the pre-commitment power. Importantly, the defendant's expected loss is maximized at higher charge levels compared to the prosecutor's payoff in the trial stage.

Within the equilibrium of the model, a prosecutor with sufficiently high pre-commitment power opts for high charges to bind herself in the trial stage and credibly threatens the defendant with severe charges. Consequently, the prosecutor places herself in a position where she incurs a loss if the trial stage is reached. In equilibrium, the trial stage is never reached.

The conducted experiment closely follows the theoretical model, utilizing two levels of commitment: high and low. With high commitment, prosecutors can inflate the charges to establish a credible commitment to severe charges. Conversely, with low commitment, prosecutors struggle to credibly commit to optimal charges, rendering any threats as mere cheap talk.

The experiment provides an ideal framework for examining subjects' behavior when offers are rejected, thereby offering insights into their responses to predicted off-path equilibrium scenarios. The first main result shows that prosecutors employ the commitment to choose higher charges than they would choose in the absence of commitment. Comparing treatments, prosecutors choose lower initial charges when commitment is high, and proportionally, they reduce charges more compared to situations with lower commitment. Nonetheless, the final charges are higher in the high-commitment treatments.

Despite binding their own actions, subjects in the experiment displayed more caution in choosing charges than predicted. Prosecutors opt for lower initial charges than anticipated in high-commitment treatments, resulting in lower defendant losses than expected and less severe expected payoffs for prosecutors if the trial stage occurs.

The second main result indicates that, although prosecutors credibly threaten defendants with higher losses when commitment is high, there is no significant difference between plea deal offers made and accepted offers. The distribution of offers remains consistent across high and low commitment levels, despite different defendant losses associated with varying levels of commitment.

Related Literature: The model presented in this paper is related to pre-commitment literature in bargaining, which starts with Schelling (1956) and Schelling (1960), who described

the bargaining process as an attempt of players to commit themselves to a position, and credible convince the others that conceding is not possible. Early formalization of Schelling's ideas are found in Crawford (1982), who presents a two-period bilateral bargaining model with concessions in the second period, and Muthoo (1992) and Muthoo (1996) who formalize the concession cost, making it proportional to the size of it. Cuellar and Rentschler (2022) present a model of adversarial bargaining in which the commitment is to a threat rather than a demand.

Other papers that include the notion of commitment to a position are: Dutta (2012), Basak and Deb (2020),Basak (2021), Dutta (2021), Miettinen and Perea (2015), Ellingsen and Miettinen (2008), Li (2011), Chung and Wood (2019), Ellingsen and Miettinen (2014), Miettinen and Vanberg (2020), Levenotoğlu and Tarar (2005), and Tarar and Leventoğlu (2009).

The experimental literature that studies commitment in bargaining is slim. Embrey et al. (2015) shows that the predictions of the reputational bargaining literature hold: if there is a probability that there exists a behavioral type that is fully committed to an offer and never changes it, the rational bargaining imitates him to gain bargaining power. Heggedal et al. (2022) shows that under the presence of outside options, the incentive to imitate the behavioral type is removed and it ensures an immediate agreement whenever two rational bargainers match.[1]

Outline: The plan of the paper is as follows. Section 2 introduces the experimental design and Section 4 the results. Section 5 concludes.

2 Theory and Equilibrium Predictions

2.1 Model

In this game, two players are involved: a prosecutor (P) and a defendant (D). The game consists of two stages: plea bargaining and trial. In the plea bargaining stage, the prosecutor offers a plea deal to the defendant. If the defendant accepts the deal, both players receive their agreed-upon payoffs. If the deal is rejected, the game moves to the trial stage, which is costly for both players. Before the payoffs are determined, the prosecutor has the opportunity to reduce the charges. Then, a lottery assigns payoffs to the players.

[1]Reputational bargaining literature starts with Abreu and Gul (2000), who present the canonical mode. Several extensions have been made, among others: Kambe (1999), Abreu and Sethi (2003), Wolitzky (2012), Atakan and Ekmekci (2014), Sanktjohanser (2020), and Ekmekci and Zhang (2021).

In specific, during the plea bargaining stage, the prosecutor first chooses the initial charges x^I to threaten the defendant. The prosecutor then offers a plea deal y to the defendant. If the defendant accepts the deal, the payoffs are $u_P = y$ and $u_D = -y$. The game ends in this case. If the plea deal is rejected, the trial stage begins.

In the trial stage, both players incur a cost of k. Before the payoffs are realized, the prosecutor has the option to reduce the charges up to a parameter α. The prosecutor chooses the final charges x^F within the range of $[\alpha x^I, x^I]$. A lottery is then conducted to assign payoffs based on the following probabilities:

$$u_D(x^F) = \begin{cases} -x^F & \text{with prob } p(x^F) \\ 0 & \text{with prob } 1 - p(x^F) \end{cases} \quad \text{and} \quad u_P(x^F) = \begin{cases} \phi x^F & \text{with prob } p(x^F) \\ -x^F & \text{with prob } 1 - p(x^F). \end{cases}$$

The probability $p(x^F)$ that the prosecutor wins at the trial stage is inversely proportional to the charges, and it is given by:

$$p(x^F) = \frac{\bar{x} - x^F}{\bar{x}}.$$

The parameter $\phi \geq 1$ represents the additional value for the prosecutor to win compared to losing at the trial stage.

We focus on the case in which the prosecutor's expected payoff $\mathbb{E}u_P(x^F)$ is lower defendant's expected loss $-\mathbb{E}u_D(x^F)$ for any $x^F \in [0, \bar{x}]$. Figure 1 provides a representation of the expected payoff and loss.

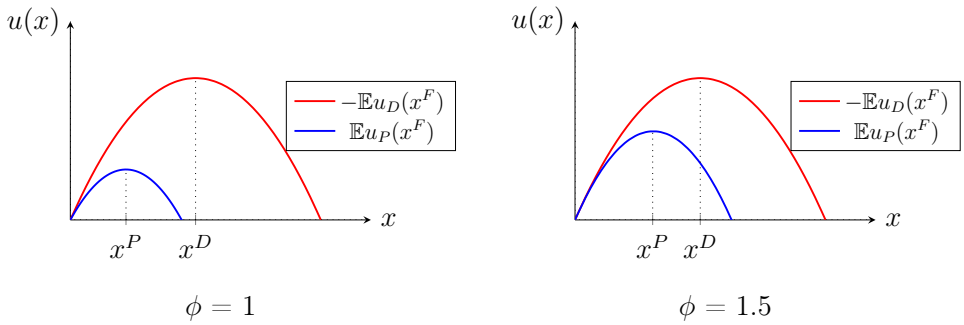


Figure 1: Equilibrium initial threat and offer.

The figure illustrates that the maximizers of the two functions are different. The maximizer x^P represents the charges chosen by the prosecutor that maximize the prosecutor's expected payoff $\mathbb{E}u_P(x)$, while x^D represents the charges that maximize the expected loss for the defendant $-\mathbb{E}u_D(x)$. The prosecutor faces a trade-off, as $x^P < x^D$. Although the

prosecutor would like to threaten the defendant with x^P, it is not sequentially rational to keep $x^F = x^P$ if the trial stage starts.

Figure 2 depicts the timeline of the game, showing the different stages: plea bargaining and trial.

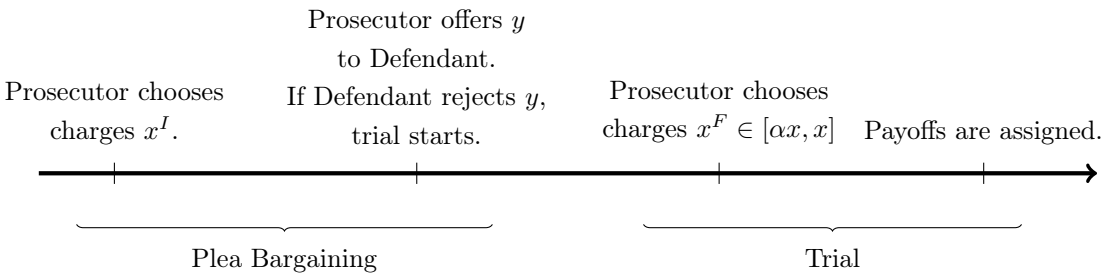


Figure 2: Timeline of the game

2.2 Equilibrium Behavior

The main insight is that a rational prosecutor chooses larger charges as the parameter α decreases. The parameter α represents the prosecutor's commitment power, as the prosecutor can only reduce the charges up to αx at the trial stage.

If the prosecutor chooses charges $x > x^P$ and the defendant rejects the plea deal, she will choose $x^F = \max x^P, \alpha x$ to maximize her expected payoff, given the self-imposed restriction $x^F \in [\alpha x, x]$. The defendant accepts the deal y if $y \leq -\mathbb{E}u_D(x^F)$. Therefore, the optimal value of the deal that the defendant accepts depends on the choice of x and the commitment level. We consider the following cutoffs:

$$\underline{\alpha} = \frac{x^P}{\bar{x}} \quad \text{and} \quad \bar{\alpha} = \frac{x^R}{\bar{x}},$$

where $\bar{\alpha} > \underline{\alpha}$.

High commitment level: If the commitment level is high ($\alpha \geq \bar{\alpha}$), the prosecutor chooses $x = \frac{x^D}{\alpha}$ and offers $y = -\mathbb{E}u_D(x^D)$, as the final charges will be $x^F = \alpha\frac{x^D}{\alpha} = x^D$ if the plea deal is rejected. In equilibrium, the defendant accepts the plea deal, and the payoffs are:

$$u_P = -\mathbb{E}u_D(x^D) \quad \text{and} \quad u_D = \mathbb{E}u_D(x^D).$$

For the experimental setting in the next section, the analysis of the off-the-equilibrium-path strategies is relevant. In that case, the prosecutor chooses $x^F = x^D$, and the defendant gets the same (expected) payoff as accepting the deal. However, the prosecutor obtains a

lower payoff, which can also be negative as shown in Figure 1. In other words, the prosecutor *ties her hands* when choosing the initial charges as a way to induce the defendant to accept a high plea deal.

Note that the equilibrium initial charges are decreasing in α. That is, the prosecutor chooses higher charges as a complement to the lack of perfect commitment power.

Intermediate commitment level: If the commitment level takes an intermediate value ($\underline{\alpha} \leq \alpha < \bar{\alpha}$), the prosecutor chooses $x = \bar{x}$ and offers $y = -\mathbb{E}u_D(\alpha\bar{x})$. Since $\alpha\bar{x} < x^D$ for any $\alpha < \bar{\alpha}$, the best the prosecutor can do is to choose $\bar{x}$. In equilibrium, the prosecutor offers $y = -\mathbb{E}u_D(\alpha\bar{x})$, and the defendant accepts it. The payoffs are given by:

$$u_P = -\mathbb{E}u_D(\alpha\bar{x}) \quad \text{and} \quad u_D = \mathbb{E}u_D(\alpha\bar{x}).$$

In off-the-equilibrium-path strategies, if the defendant rejects the offer, they receive the same expected payoff as the deal, and the prosecutor obtains a lower payoff. It is important to note that in this case, the prosecutor's expected payoff decreases as the commitment level α decreases. However, the prosecutor still partially restricts their options, albeit to a lesser extent than if α were high.

Low commitment level: If the commitment level is low ($\alpha < \underline{\alpha}$), the prosecutor can choose any x in the range $[x^P, \bar{x}]$ and offers $y = -\mathbb{E}u_D(x^P)$. In this scenario, $\alpha\bar{x} < x^P$, meaning that even by selecting the highest possible charges, the prosecutor will opt for the final charges that maximize her expected payoff rather than maximizing the defendant's loss. Choosing $x < x^P$ is not optimal for the prosecutor since she will want to increase the charges during the trial stage.

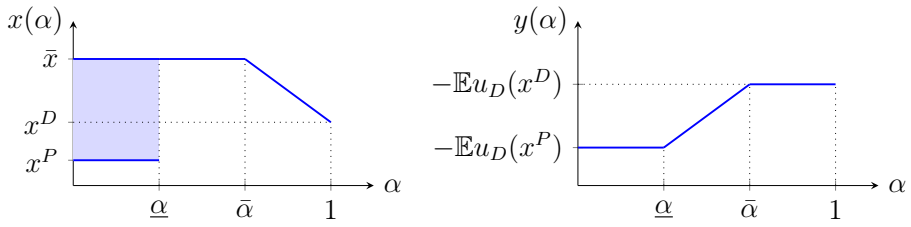
In equilibrium, the prosecutor offers $y = -\mathbb{E}u_D(x^P)$, and the defendant accepts the plea deal. However, if the defendant rejects the plea deal, the prosecutor chooses $x^F = x^P$ and obtains her highest expected payoff during the trial stage.

In this situation, we refer to the prosecutor as engaging in *bluffing* if she chooses $x > x^P$ since she cannot sustain those charges in equilibrium.

Summary: In equilibrium, the election of the initial charges and the optimal deal to be offered are as follows:

$$x(\alpha) = \begin{cases} [x^P, \bar{x}] & \text{if } \alpha \in [0, \underline{\alpha}] \\ \bar{x} & \text{if } \alpha \in (\underline{\alpha}, \bar{\alpha}] \\ \frac{x^R}{\alpha} & \text{if } \alpha \in (\bar{\alpha}, 1] \end{cases} \quad \text{and} \quad y(\alpha) = \begin{cases} -\mathbb{E}u_D(x^P) & \text{if } \alpha \in [0, \tilde{\alpha}] \\ -\mathbb{E}u_D(\alpha\bar{x}) & \text{if } \alpha \in (\tilde{\alpha}, \bar{\alpha}] \\ -\mathbb{E}u_D(x^D) & \text{if } \alpha \in (\bar{\alpha}, 1], \end{cases}$$

and the defendant accepts the offer. Figure 3 graphically shows the initial charges and payoffs as a function of the commitment level.



(a) Initial charges (b) Optimal deal

Figure 3: Equilibrium initial charges and optimal plea deal.

The equilibrium predictions assume that the players follow the equilibrium path, where the trial stage is never reached. However, in the event of a deviation from the equilibrium path, specifically, when the defendant rejects the deal y, the prosecutor's choice of final charges is determined as follows: if $x \leq x^P$, the final charges are set to $x^F = x$, while if $x > x^P$, the prosecutor selects $x^F = \max x^P, \alpha x$.

3 Experimental Design

The experiment was designed based on the game described in Section 2, utilizing parameters $\bar{x} = 100$ and $k = 10$. The function $p(x^F)$ was defined as:

$$p(x^F) = 1 - \frac{x^F}{100}.$$

Four treatments were conducted, each representing a combination of a level of ϕ and a level of α. Specifically, ϕ was chosen from the set $1, 1.5$, while α was chosen from the set $0.2, 0.7$. The four treatments were denoted as T1, T2, B3, and B4, as indicated in the following table.

Treatment	α	ϕ
T1	0.7	1
T2	0.7	1.5
B3	0.2	1
B4	0.2	1.5

We denote "truthful treatments" to treatments T1 and T2, while we refer to B3 and B4 as "bluffing treatments." This categorization is based on whether the prosecutor possessed sufficient commitment power to credibly threaten the defendant with maximizing their loss. The main focus of the paper is to examine the impact of commitment by comparing pairwise between T1-B3 and T2-B4, where the only difference between them was the value of α. By considering two values of ϕ, the aim is to assess if the effect of commitment on behavior remained consistent in the presence of a more attractive payoff at the trial stage.

The experiment took place at Utah State University and involved undergraduate students from various majors. It employed a within-subjects design for α and a between-subjects design for ϕ. The experiment comprised 12 sessions, with each session accommodating 10 participants. In each session, there were 30 rounds, with 15 rounds featuring $\alpha = 0.7$ and the remaining 15 rounds featuring $\alpha = 0.2$. Half of the sessions maintained a constant $\phi = 1$ throughout all rounds, while the other half utilized $\phi = 1.5$ for all rounds.

Each round involved randomly pairing two participants, with one designated as the prosecutor and the other as the defendant. The prosecutor initiated the round by selecting an initial threat $x_I \in [0, 100]$ and subsequently proposing an offer $x \in [0, 100]$. The interface displayed potential final threats if the defendant rejected the offer.[2] After the offer was made, the prosecutor had to decide whether to accept or reject it. The defendant's screen presented the initial position chosen by the prosecutor, allowing the defendant to explore possible final threats and their associated probabilities in the event of rejecting the offer. The round concluded either when the defendant accepted the offer or when the prosecutor determined the final threat upon offer rejection, with the payoffs being realized accordingly.

3.1 Key equilibrium predictions

The first key prediction is that the prosecutor leverages her commitment power to strategically limit her options and be forced to choose high final charges. This implies that the initial charges must align with her ability to commit to selecting a final charge of x^D when $\alpha = 0.7$ (where $0.7 > \bar{\alpha}$).

[2]Refer to Appendix A for screenshots.

The second prediction focuses on how the prosecutor utilizes this credible threat to induce the defendant to accept a high plea deal in the case of $\alpha = 0.7$. This prediction comprises two parts: firstly, the prosecutor will make an appropriate offer that surpasses the offer for $\alpha = 0.2$, and secondly, the defendant will accept that offer.

The theoretical predictions for each variable are summarized in Table 1.

	T1 (0.7,1)	T2 (0.7,1.5)	B3 (0.2,1)	B4 (0.2,1.5)
Initial charges	71.4	71.4	[25,100]	[30,100]
Offer	35	35	28.7	31
Percentage accepted	1	1	1	1
Final charges	50	50	25	30
Expected payoff prosecutor	35	35	28.7	31
Expected payoff prosecutor (off-path)	-10	2.5	2.5	12.5
Expected payoff defendant	-35	-35	-28.7	-31

Table 1: Theoretical predictions.

4 Results

To mitigate potential learning effects, we exclude the first five rounds of each treatment in each session. Table 2 provides an overview of the bargaining outcomes. More than 50% of the negotiations conclude at the trial stage, which contradicts the theoretical expectations. However, this allows for an analysis of the realization of the final charges. We assess the subjects' performance by focusing on two main results. The first pertains to the utilization of commitment and hand-tying behavior, while the second examines how this behavior manifests in reaching an agreement.

	T1 (0.7,1)	T2 (0.7,1.5)	B3 (0.2,1)	B4 (0.2,1.5)
Mean initial charges	44.8	49.2	65.1	63.8
Mean offer	35.8	39.8	38.1	38.8
Proportion of accepted offers	0.45	0.48	0.44	0.45
Mean final charges	38.3	41	33.9	32.4
Prosecutor's mean expected payoff (predicted)	-1.1	8.1	2.3	12.0
Prosecutor's mean expected payoff (realized)	-5.3	4.5	-2.9	8.8
Defendant's mean expected payoff (predicted)	-30.8	-31.7	-28.3	-30.4
Defendant's mean expected payoff (realized)	-31.5	-32.2	-30.5	-30.5

Table 2: Plea bargaining outcomes.

Table 2 presents the expected payoffs for both the prosecutor and the defendant. The *predicted* values represent the expected payoffs calculated based on the assumption that the prosecutor chooses the optimal final charges, given their chosen initial charges. On the other hand, the *realized* values indicate the expected payoffs at the trial stage for those players whose offers were rejected, and these values are calculated using their actual chosen final charges.

4.1 Effect of the commitment on the threat

Result 1: *The prosecutor strategically employs her commitment power to tie her hands and ensure final charges higher than* x^P.

The distribution of initial threats across the four treatments is illustrated in Figure 4, showcasing a wide range of values. The average initial charges chosen by the prosecutor in treatments T1 and T2 are lower compared to treatments B3 and B4. Specifically, the mean initial charges for T1 and T2 are 44.8 and 49.2, respectively, while for B3 and B4, they are 65.1 and 63.8, respectively. Figure 4 reveals that the distribution of initial threats aligns with subjects who comprehend the significance of commitment. In treatments T1 and T2, only 1.7% and 2.3% of initial threats are higher or equal to 95, respectively, while for B3 and B4, these percentages rise to 25% and 15%, respectively.

Definition: The scaling down index for subject i is defined as:

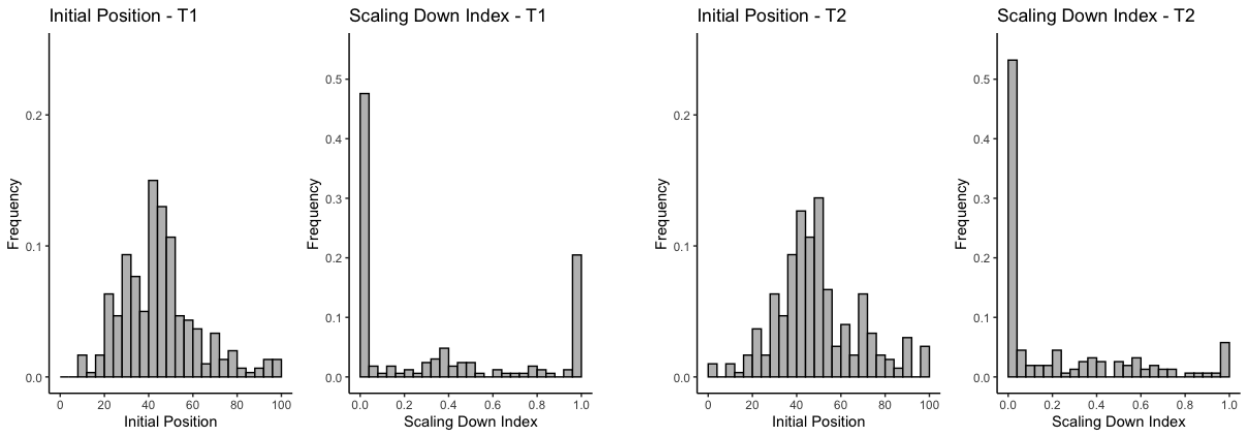
$$SDI_i = \frac{x_i^F - \alpha x_i^I}{(1-\alpha)x_i^I}.$$

The scaling down index ranges from 0 to 1, where an SDI of 1 indicates that the prosecutor did not reduce the charges, meaning $x_F = x_I$. Conversely, an SDI of 0 implies that the prosecutor reduced the charges as much as possible, resulting in $x_F = \alpha x_I$.

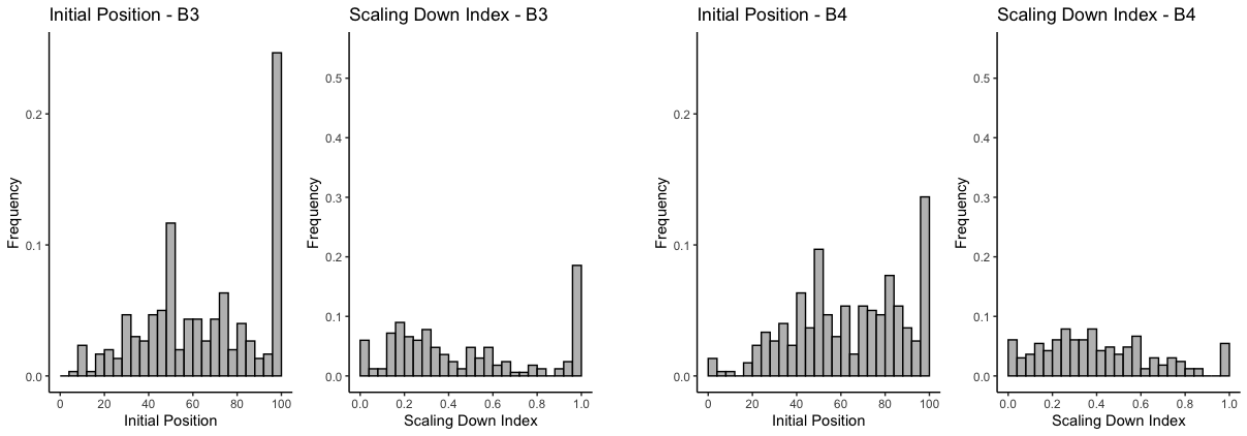
For prosecutors whose offers were rejected, the average scaling down index in treatments T1 and T2 is 0.35 and 0.22, respectively, while in treatments B3 and B4, it is 0.48 and 0.41, respectively. Comparing T1 and T2 with B3 and B4, it is evident that the prosecutors scaled down the threat more in T1 and T2 [p-value = 2.686e-06 and 1.234e-12]. Figure 4 depicts the difference in the distribution of the scaling down index between treatments T and B. The proportion of scaling down index values between 0 and 0.04 is 48% and 53% for T1 and T2, while for B3 and B4, it is 6% and 6.1%, respectively.

This behavior aligns with the election of initial charges aimed at tying hands regarding the final charges. In treatments T1 and T2, the prosecutor, on average, chose lower initial

charges compared to B3 and B4 and reduced them to a greater extent. Consequently, the mean final charges for T1 and T2 are higher (38.3 and 41) compared to B3 and B4 (33.9 and 32.4).



(a) Treatment T1 **(b)** Treatment T2



(c) Treatment B3 **(d)** Treatment B4

Figure 4: Initial charges (position) and scaling down index if the offer were rejected.

Result 2: *The prosecutor engages in bluffing behavior in treatments B3 and B4.*

According to theoretical predictions, any initial charges higher than 25 and 30 are considered an equilibrium for treatments B3 and B4, respectively. An interesting aspect concerning multiple equilibria is the selection of which equilibrium to pursue. The experimental results demonstrate that the prosecutor opts for initial charges of 65.1 and 63.8 in B3 and B4, respectively, which are not significantly different [p-value = 0.4929]. Additionally, as mentioned earlier, 15% and 25% of the initial charges exceed 95. This suggests that subjects strategically choose large initial charges (greater than in T1 and T2) as a means of bluffing about the likelihood of imposing high final charges.

Result 3: *The prosecutor exhibits less hand-tying behavior than predicted theoretically.*

In treatments T1 and T2, the prosecutor does engage in hand-tying by selecting an appropriate x_I. However, the degree of hand-tying observed in the experiment is lower than what was theoretically predicted. According to theory, the prosecutor should choose initial charges of 71 and reduce them to 50 if the offer is rejected. However, the experiment yields lower values, with initial charges of 44.8 and 49.2 for T1 and T2, respectively, and final charges of 38.3 and 41.

These results suggest a tendency towards risk aversion when facing a rejected offer, as the predicted expected payoff (conditional on their x_I) is higher than the theoretical prediction. The predicted expected payoff of -1.1 and 8.1 is significantly higher than the predicted values of -10 and 2.5. Furthermore, the realized expected values, taking into account the final charges of players whose offers were rejected, are also higher than the theoretical predictions: -5.3 and 4.5. However, they remain lower than the predicted expected payoff conditional on x_I. This indicates that the defendant's expected loss (both predicted and realized) is lower than what was theoretically predicted.

Conversely, in treatments B3 and B4, the final threats are higher than what was theoretically predicted: 33.9 and 32.4 compared to 25 and 30. As a result, the realized expected payoff for the prosecutor is lower than the theoretical prediction: -2.9 and 8.8 for B3 and B4, respectively, while the theoretical predictions are 2.5 and 12.5.

4.2 Effect of the commitment on the agreement

Result 4: *The acceptance rate of the offer is low primarily due to the prosecutor making higher-than-optimal offers.*

The percentage of accepted offers is consistently lower than the theoretical prediction of 100%, as shown in Table 2. Across all treatments, the acceptance rate is approximately 45%, and there is no significant difference between treatments. Table 3 provides evidence that the high offers made by the prosecutor are the main driver behind the low acceptance rate.

In each treatment, the mean offer made by the prosecutor exceeds the mean expected loss for the defendant. Less than 45% of the offers are lower or equal to the defendant's predicted expected loss. Conversely, the defendant's acceptance rate is around 75% for offers that are lower or equal to their predicted expected loss. The acceptance rate declines for offers that exceed the predicted expected loss, although there is no significant difference observed among the treatments.

	T1 (0.7,1)	T2 (0.7,1.5)	B3 (0.2,1)	B4 (0.2,1.5)
Mean defendant's expected loss	30.8	31.7	28.3	30.4
Mean offer	35.8	39.8	38.1	38.8
Mean accepted offer	23.8	29.7	26.5	26.5
Mean rejected offer	45.5	49.2	47.3	48.9
% offer (Offer $\leq$ $\mathbb{E}$Loss)	0.45	0.36	0.34	0.40
% offer (Offer $\leq 1.25 \times \mathbb{E}$Loss)	0.68	0.58	0.61	0.57
% offer (Offer $\leq 1.5 \times \mathbb{E}$Loss)	0.80	0.79	0.75	0.77
% accepted (Offer $\leq$ $\mathbb{E}$Loss)	0.74	0.81	0.72	0.77
% accepted (Offer $\leq 1.25 \times \mathbb{E}$Loss)	0.62	0.71	0.61	0.64
% accepted (Offer $\leq 1.5 \times \mathbb{E}$Loss)	0.55	0.58	0.56	0.58

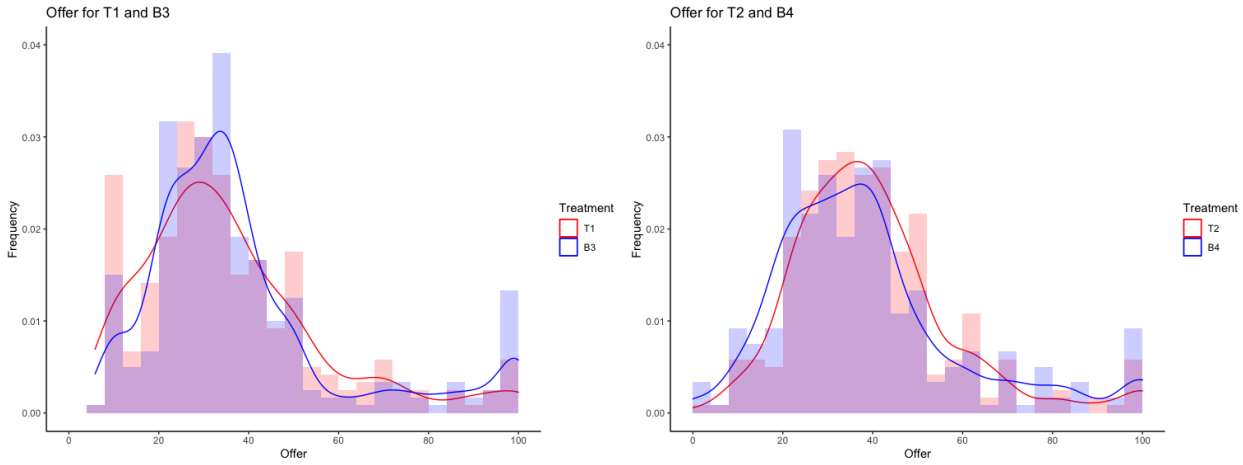
Table 3: Analysis of the offers made and accepted.

Result 5: *There is no significant effect of committing to higher final charges on choosing the offer and deciding whether to accept it.*

There is no statistically significant difference observed in the mean offers made between treatment T1 and B3 (35.8 and 38.1) and between T2 and B4 (39.8 and 38.8). Similarly, there is no significant difference in the mean accepted offer between T1 and B3 (23.8 and 26.5) and T2 and B4 (29.7 and 26.5). The distribution of offers is also similar, as depicted in Figure 5.

The theoretical prediction suggests that the prosecutor should make higher offers in treatments T1 and T2. However, the experimental results indicate that the prosecutor chooses lower initial charges than predicted for T1 and T2, leading to lower final charges as well. Conversely, in treatments B3 and B4, if an offer is rejected, the prosecutor chooses higher final charges than theoretically predicted. These findings imply that the expected losses for the defendant in treatments P and B are closer than anticipated, although they are still statistically different.

The estimated expected losses for the defendant in T1 and B3 are -30.8 and -28.3, respectively, and they are significantly different. However, the realized expected payoffs are -31.5 and -30.5, which are significantly different at a 5% level but not at a 1% level (p-value = 0.038). For treatments T2 and B4, the estimated expected payoffs are -31.7 and -30.4, while the realized expected payoffs are -32.2 and -30.5, respectively. In both cases, the realized expected payoffs are significantly different. Thus, it is unlikely that the prosecutor making the same offer for T1-B3 and T2-B4 is solely driven by the same expected payoff.



(a) Offers for Treatment T1 and B3

(b) Offers for Treatment T2 and B4

Figure 5: Distribution of offers.

Whether to accept the offer does not exhibit any significant difference between treatments T1-B3 and T2-B4. Table 4 presents the proportion of accepted offers within the range of 20 to 50, which accounts for 63%, 73%, 72%, and 66% of the total offers for treatments T1, T2, B3, and B4, respectively. The proportions of accepted offers per interval for T1 and B3, as well as T2 and B4, do not show any significant difference, except for offers within the interval of $[30, 40)$ for treatments T2 and B4. In this particular interval, the difference is significant at a 5% confidence level but not at a 1% level (p-value = 0.01007).

	T1 (0.7,1)	T2 (0.7,1.5)	B3 (0.2,1)	B4 (0.2,1.5)
Number of offers in [20,30)	80	71	90	77
% accepted offers in [20,30)	0.68	0.82	0.57	0.74
Number of offers in [30,40)	66	79	86	72
% accepted offers in [30,40)	0.38	0.56	0.48	0.35
Number of offers in [40,50)	47	70	41	53
% accepted offers in [40,50)	0.13	0.24	0.22	0.38

Table 4: Distribution of offers made between 20 and 50.

The results of the Logit regression show that the offer itself is the only significant parameter in the acceptance decision, while the expected payoffs do not show any significant influence, as presented in Table 5.

	T1	T2	B3	B4
(Intercept)	5.81	4.24	8.66	−5.85
	(4.50)	(3.97)	(9.24)	(6.12)
Offer	−0.13***	−0.13***	−0.10***	−0.12***
	(0.02)	(0.02)	(0.01)	(0.02)
Initial threat	0.10	0.02	0.01*	0.00
	(0.06)	(0.05)	(0.01)	(0.01)
Round	0.03	0.08	0.01	−0.03
	(0.05)	(0.05)	(0.05)	(0.05)
Order	−0.39	−0.55	−0.35	−0.96
	(0.83)	(0.82)	(0.78)	(0.82)
defendant's expected payoff	0.21	0.05	0.26	−0.60
	(0.22)	(0.20)	(0.39)	(0.33)
prosecutor's expected payoff	0.25*	0.06	0.49	−0.55
	(0.10)	(0.07)	(0.83)	(0.38)
Num. obs.	300	300	300	300

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 5: Logit regressions.

5 Conclusion

The experimental results reveal interesting patterns in the behavior of subjects acting as prosecutors in different treatments. For treatments T1 and T2, subjects tend to tie their hands by choosing large initial charges, although the values are significantly lower than the theoretically predicted ones. This behavior suggests a level of risk aversion in reducing the charges if the offer is rejected. On the other hand, in treatments B3 and B4, subjects bluff by choosing large initial charges as a way to signal high final charges.

Notably, the mean offers made by the prosecutors are significantly higher than their expected payoffs, which were optimally calculated based on the initial charges. The acceptance rate of offers is relatively low (around 40%) due to the high offers made by the prosecutor. However, when the offers are within the range of the expected loss for the defendant, the acceptance rate increases to over 70%.

Interestingly, the final charges for cases where the offer was rejected tend to be higher than the optimal final charges, although it is still centered around the predicted value. The realized expected payoffs for the prosecutor are significantly higher than the theoretical predictions for treatments T1 and T2, but lower for treatments B3 and B4. This indicates that the subjects' behavior deviates from the theoretical predictions but still follows the

general trend.

The observed discrepancies between the experimental results and the theoretical predictions suggest that subjects do not fully utilize the commitment power to tie their hands, but rather exhibit partial use of this strategy. This suggests that the parameter ϕ plays a significant role in influencing their behavior.

Overall, this study provides insights into the complex dynamics of negotiation and the interplay between commitment, bluffing, and risk aversion in the decision-making process of prosecutors.

References

Abreu, D. and Gul, F. (2000). Bargaining and reputation. *Econometrica*, 68(1):85–117.

Abreu, D. and Sethi, R. (2003). Evolutionary stability in a reputational model of bargaining. *Games and Economic Behavior*, 44(2):195–216.

Atakan, A. E. and Ekmekci, M. (2014). Bargaining and reputation in search markets. *Review of Economic studies*, 81(1):1–29.

Basak, D. (2021). Political bargaining under incomplete information about public reaction.

Basak, D. and Deb, J. (2020). Gambling over public opinion. *American Economic Review*, 110(11):3492–3521.

Chung, B. W. and Wood, D. H. (2019). Threats and promises in bargaining. *Journal of Economic Behavior & Organization*, 165:37–50.

Crawford, V. P. (1982). A theory of disagreement in bargaining. *Econometrica: Journal of the Econometric Society*, pages 607–637.

Cuellar, P. and Rentschler, L. (2022). Threat, commitment and brinkmanship in adversarial bargaining. *Working Paper*.

Dutta, R. (2012). Bargaining with revoking costs. *Games and Economic Behavior*, 74(1):144–153.

Dutta, R. (2021). Bargaining as a struggle between competing attempts at commitment. *Available at SSRN 3800521*.

Ekmekci, M. and Zhang, H. (2021). Reputational bargaining with ultimatum opportunities. *arXiv preprint arXiv:2105.01581*.

Ellingsen, T. and Miettinen, T. (2008). Commitment and conflict in bilateral bargaining. *American Economic Review*, 98(4):1629–35.

Ellingsen, T. and Miettinen, T. (2014). Tough negotiations: Bilateral bargaining with durable commitments. *Games and Economic Behavior*, 87:353–366.

Embrey, M., Fréchette, G. R., and Lehrer, S. F. (2015). Bargaining and reputation: An experiment on bargaining in the presence of behavioural types. *The Review of Economic Studies*, 82(2):608–631.

Heggedal, T.-R., Helland, L., and Knutsen, M. V. (2022). The power of outside options in the presence of obstinate types. *Games and Economic Behavior*.

Kambe, S. (1999). Bargaining with imperfect commitment. *Games and Economic Behavior*, 28(2):217–237.

Levenotoğlu, B. and Tarar, A. (2005). Prenegotiation public commitment in domestic and international bargaining. *American Political Science Review*, 99(3):419–433.

Li, D. (2011). Commitment and compromise in bargaining. *Journal of Economic Behavior & Organization*, 77(2):203–211.

Miettinen, T. and Perea, A. (2015). Commitment in alternating offers bargaining. *Mathematical Social Sciences*, 76:12–18.

Miettinen, T. and Vanberg, C. (2020). Commitment and conflict in multilateral bargaining.

Muthoo, A. (1992). Revocable commitment and sequential bargaining. *The Economic Journal*, 102(411):378–387.

Muthoo, A. (1996). A bargaining model based on the commitment tactic. *Journal of Economic theory*, 69(1):134–152.

Sanktjohanser, A. (2020). Optimally stubborn.

Schelling, T. (1960). The strategy of conflict harvard university press cambridge.

Schelling, T. C. (1956). An essay on bargaining. *The American Economic Review*, 46(3):281–306.

Tarar, A. and Leventoğlu, B. (2009). Public commitment in crisis bargaining. *International Studies Quarterly*, 53(3):817–839.

Wolitzky, A. (2012). Reputational bargaining with minimal knowledge of rationality. *Econometrica*, 80(5):2047–2087.

Appendices

A Experimental Instructions

The following boxes show the experiment instructions and decision-making screens that were seen by subjects in the treatment configuration with values $\phi = 1.5$ and $\alpha = 0.2$ for the first 15 rounds, and $\alpha = 0.7$ for the subsequent 15 rounds. Instructions for other treatment configurations are omitted to avoid redundancy but can be provided by the authors upon request.

Introduction

Welcome to this experiment. The decisions you make during this experiment will determine how much money you earn. You will be paid in cash, privately, at the end of our experiment.

It is important that you remain silent and do not look at other people's work. If you have any questions, or need assistance of any kind, please raise your hand and an experimenter will come to you. If you talk, laugh, exclaim out loud, etc., you will be asked to leave and you will not be paid. We expect and appreciate your cooperation.

For today's experiment, you will receive an initial payment of \$18, in addition to the \$7 show-up fee. Your earnings in the experiment are expressed in Experimental Francs (EF). 5 EF are worth \$1. At the end of the experiment, your total earnings will be paid to you in cash, privately.

The experiment consists of two parts, labeled Part 1 and Part 2. Each part has 15 decision rounds. We will read you the instructions for Part 1 now. After completing Part 1 we will read instructions for Part 2. After we have read the instructions, there will be time to ask clarifying questions. When we are done going through the instructions for Part 1, each of you will have to answer a few brief questions to ensure everyone understands.

Groups and roles

At the beginning of each round, all participants will be randomly divided into groups of two. You will only interact with the other participant randomly assigned to your group through the computer interface. You will not know who the other member of your group is. Within each round you will have no interaction with participants who are randomly assigned to other groups.

Groups are randomly determined every round, so that the other participant in your group is probably not the same between rounds.

In each round, one participant in each group will be the PROPOSER, and the other participant will be the RESPONDER. These roles are randomly determined in each group in each round, so that you will probably not be in the same role in all rounds.

Round overview

During each round, an outcome is determined. There are two possible outcomes: OUTCOME P and OUTCOME R.

Each round is comprised of several stages.

Stage 1: The PROPOSER chooses an INITIAL POSITION, which is a number between 0 and 100, inclusive.

Stage 2: The PROPOSER chooses an OFFER, which is also a number between 0 and 100, inclusive.

Stage 3: The RESPONDER sees the INITIAL POSITION and the OFFER. After doing so, the RESPONDER decides whether to accept or reject the OFFER. If the RESPONDER rejects the OFFER, then the round earnings of the RESPONDER and the PROPOSER are both reduced by 10. If the RESPONDER accepts the OFFER, the round ends, and the outcome is OUTCOME P.

The round earnings are:

PROPOSER: OFFER

RESPONDER: $(-1)\cdot$ (OFFER)

Stage 4: If the RESPONDER rejects the OFFER, the PROPOSER chooses a FINAL POSITION. The FINAL POSITION is any number between $0.2\cdot$ (INITIAL POSITION) and the INITIAL POSITION, inclusive. That is, when choosing the FINAL POSITION, the PROPOSER can reduce their INITIAL POSITION by no more than 80%. The FINAL POSITION cannot be higher than the INITIAL POSITION.

Stage 5: If the RESPONDER rejects the OFFER, the outcome is determined using the FINAL POSITION chosen by the PROPOSER. The outcome will be OUTCOME R with probability (FINAL POSITION)%, and will be OUTCOME P with probability (100 − FINAL POSITION)%. Note that the higher the FINAL POSITION, the lower the probability that the outcome is OUTCOME P.

If the outcome is OUTCOME R, round earnings are

PROPOSER: $(-1)\cdot$ (FINAL POSITION) $-\ 10$

RESPONDER: $0 - 10 = -10$

If the outcome is OUTCOME P, round earnings are

PROPOSER: $(1.5)\cdot$ (FINAL POSITION) $-\ 10$

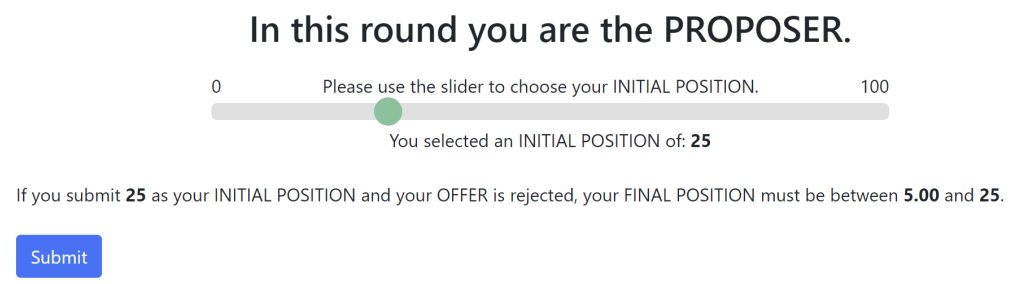
RESPONDER: $(-1)\cdot$ (FINAL POSITION) $-\ 10$

PROPOSER chooses their INITIAL POSITION

At the beginning of each round the PROPOSER chooses an INITIAL POSITION, which can be any number between 0 and 100, inclusive. The INITIAL POSITION is chosen by moving the slider to the desired number and clicking "Submit".

When the PROPOSER moves the slider, the text below the slider will indicate the FINAL POSITIONs that could be chosen if the current position of the slider were submitted as the INITIAL POSITION. Recall that when choosing the FINAL POSITION, the PROPOSER can reduce their INITIAL POSITION by no more than 80%, and the FINAL POSITION cannot be higher than the INITIAL POSITION.

Note that the starting location of the slider is chosen randomly.



PROPOSER chooses their OFFER

After submitting the INITIAL POSITION, the PROPOSER chooses an OFFER. The OFFER can be any number between 0 and 100, inclusive. The OFFER is chosen by moving the slider to the desired number and clicking "Submit".

When choosing the OFFER, the PROPOSER's screen will display:

1. Their INITIAL POSITION
2. The FINAL POSITIONs that they can choose if their OFFER is rejected.
3. A reminder that if the OFFER is rejected by the RESPONDER, the round earnings of the PROPOSER and the RESPONDER will decrease by 10.
4. A second slider whose end points correspond to the possible FINAL POSITIONs the PROPOSER can choose if the OFFER is rejected. The PROPOSER can use this slider to explore scenarios corresponding to possible FINAL POSITIONS. For any possible FINAL POSITION, a bar chart will illustrate the probability that the outcome will be OUTCOME R or OUTCOME P, and will state the round earnings that both the PROPOSER and RESPONDER would receive.

Note that the starting location of the slider is set to the previously chosen INITIAL POSITION.

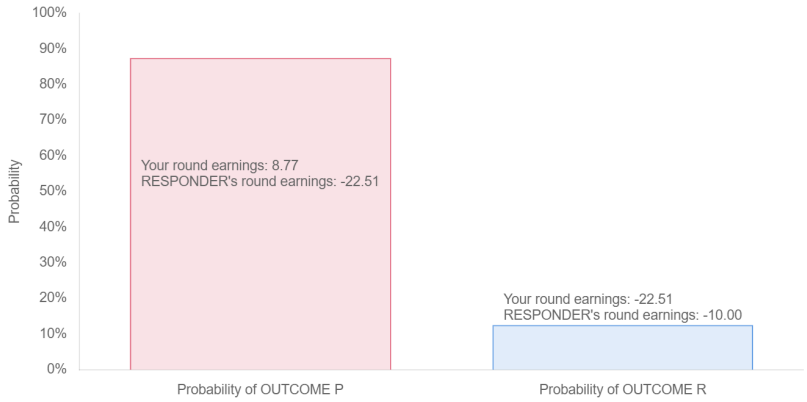
Your INITIAL POSITION is 25.0

If your OFFER is rejected, your FINAL POSITION must be between **5.0** and **25.0**.
If the RESPONDER rejects your OFFER your round earnings and the round earnings of the RESPONDER will be reduced **10**

---

0 Please use the slider to choose your OFFER. 100

You can use the slider below to explore possible outcomes if the RESPONDER rejects your offer.



FINAL POSITIONS you can choose if your OFFER is rejected.

5.0 Potential FINAL POSITION: **12.51** **25.0**

RESPONDER accepts or rejects the OFFER

After the PROPOSER chooses the INITIAL POSITION and the OFFER, the RESPONDER sees them, and decides whether to accept or reject the OFFER.

The RESPONDER accepts the OFFER by clicking "Accept" and rejects the OFFER by clicking "Reject".

Next to the "Accept" button, the RESPONDER's screen will state that if they accept the OFFER the outcome will be OUTCOME P, their round earnings will be equal to $(-1) \cdot$ OFFER, and the PROPOSER's round earnings will be equal to the OFFER.

Next to the "Reject" button, the RESPONDER's screen will state that if they reject the OFFER, both their round earnings and the round earnings of the PROPOSER will decrease by 10. The screen will also indicate the FINAL POSITIONS that the PROPOSER can choose if the OFFER is rejected. (Recall that when choosing the FINAL POSITION, the PROPOSER can reduce their INITIAL POSITION by no more than 80%, and the FINAL POSITION cannot be higher than the INITIAL POSITION.).

In addition, the RESPONDER's screen will display a slider whose end points correspond to the possible FINAL POSITIONs the PROPOSER can choose if the OFFER is rejected. The RESPONDER can use this slider to explore scenarios corresponding to possible FINAL POSITIONS. For any possible FINAL POSITION, a bar chart will illustrate the probability that the outcome will be OUTCOME R or OUTCOME P, and will state the round earnings that both the PROPOSER and RESPONDER would receive.

In this round you are the RESPONDER

The PROPOSER chose an INITIAL POSITION of **25.0**
The PROPOSER chose an OFFER of **23.84**

If you accept

The round ends.
The outcome would be OUTCOME P.
Your round earnings would be: **-23.84**
PROPOSER round earnings would be: **23.84**

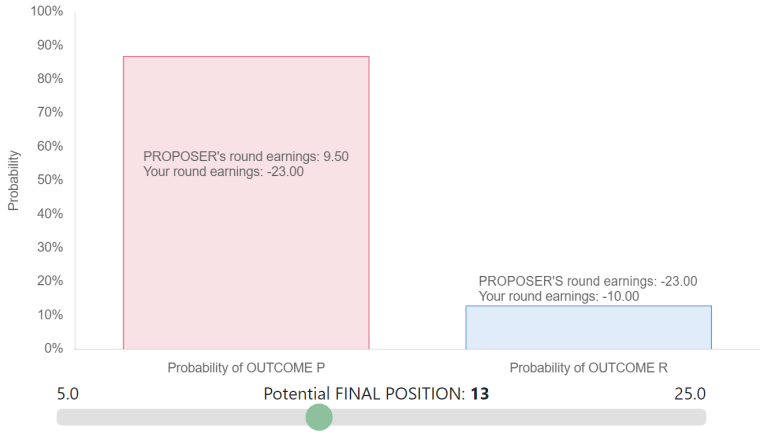
Accept

If you reject

Your round earnings and the round earnings of the PROPOSER will decrease by **10**
The PROPOSER will chose a FINAL POSITION between **5.0** and **25.0**

Reject

You may use the slider below to explore possible outcomes which will be determined by the PROPOSER's choice of FINAL POSITION.



PROPOSER chooses their FINAL POSITION

If the RESPONDER rejects the OFFER the round earnings of the RESPONDER and the PROPOSER decrease by 10. The PROPOSER then chooses a FINAL POSITION. When choosing the FINAL POSITION, the PROPOSER can reduce their INITIAL POSITION by no more than 80%. The FINAL POSITION cannot be higher than the INITIAL POSITION.

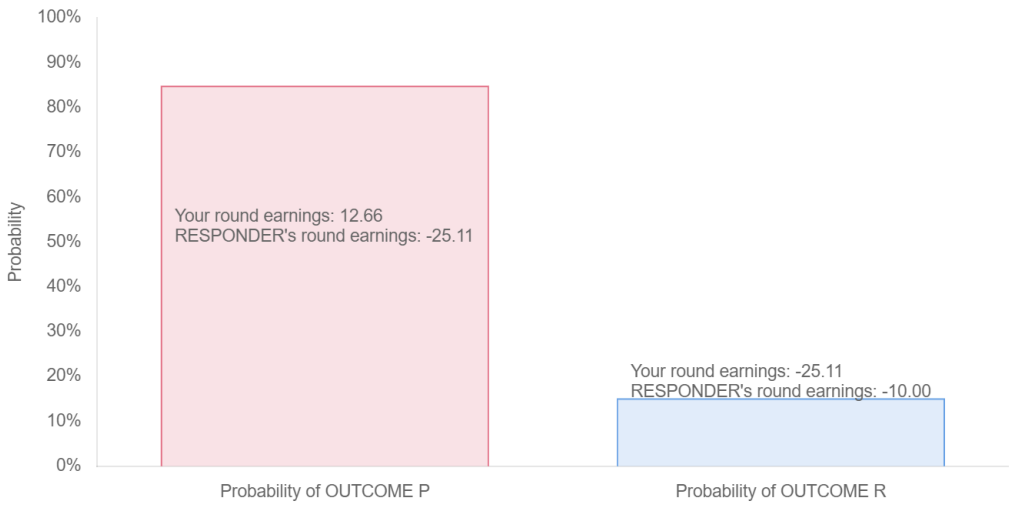
The FINAL POSITION is chosen by moving the slider to the desired number and clicking "Submit".

When the PROPOSER moves the slider, a bar chart above the slider will indicate the probability that the outcome will be OUTCOME P and the probability that the outcome will be OUTCOME R if the current position of the slider were submitted as their FINAL POSITION. This bar chart will also state the the round earnings that both the PROPOSER and RESPONDER would receive in each possible outcome if the current position of the slider were submitted as the FINAL POSITION.

Note that the starting location of the slider is chosen randomly.

The RESPONDER rejected your offer.

Your FINAL POSITION must be between **5.0** and **25.0**.



Please use the slider to choose your FINAL POSITION.

5.0 FINAL POSITION: **15.11** **25.0**

Outcome when the OFFER is rejected

If the RESPONDER rejects the OFFER, the outcome is determined using the FINAL POSITION chosen by the PROPOSER. The outcome will be OUTCOME R with probability (FINAL POSITION)%, and will be OUTCOME P with probability (100 − FINAL POSITION)%.

If the outcome is OUTCOME R, round earnings are

PROPOSER: $(-1) \cdot (\text{FINAL POSITION}) - 10$

RESPONDER: $0 - 10 = -10$

If the outcome is OUTCOME P, round earnings are

PROPOSER: $(1.5) \cdot (\text{FINAL POSITION}) - 10$

RESPONDER: $(-1) \cdot (\text{FINAL POSITION}) - 10$

Example 1

The PROPOSER chooses an INITIAL POSITION of 45, and an OFFER of 30.

The RESPONDER rejects the OFFER, so the round earnings of both the RESPONDER and the PROPOSER are decreased by 10. The PROPOSER chooses a FINAL POSITION. This FINAL POSITION must be between $0.2 \cdot 45$ and 45. That is, the FINAL POSITION must be between 9 and 45.

The PROPOSER chooses a FINAL POSITION of 40. This means that there is a 40% probability that the outcome will be OUTCOME R, and a 60% probability the outcome will be OUTCOME P.

When the outcome is resolved, it is OUTCOME R. Round earnings are

PROPOSER: $-40 - 10 = -50$

RESPONDER: $0 - 10 = -10$

Example 2

The PROPOSER chooses an INITIAL POSITION of 40, and an OFFER of 25.

The RESPONDER accepts the OFFER, so the round ends and the outcome is OUTCOME P. Round earnings are:

PROPOSER: 25

RESPONDER: -25

Example 3

The PROPOSER chooses an INITIAL POSITION of 60, and an OFFER of 35.

The RESPONDER rejects the OFFER, so the round earnings of both the RESPONDER and the PROPOSER are decreased by 10. The PROPOSER chooses a FINAL POSITION. This FINAL POSITION must be between $0.2 \cdot 60$ and 60. That is, the FINAL POSITION must be between 12 and 60.

The PROPOSER chooses a FINAL POSITION of 50. This means that there is a 50% probability that the outcome will be OUTCOME R, and a 50% probability the outcome will be OUTCOME P. When the outcome is resolved, it is OUTCOME P. Round earnings are

PROPOSER: $(1.5) \cdot (50) - 10 = 65$

RESPONDER: $-50 - 10 = -60)$

Results of a round

At the end of each round the results of the round will be displayed on the screen. The results you will see are:

1. Your role in the round.
2. The INITIAL POSITION chosen by the PROPOSER.
3. The OFFER chosen by the PROPOSER.
4. Whether the OFFER was accepted or rejected by the RESPONDER.
5. If relevant, the FINAL POSITION chosen by the PROPOSER.
6. The outcome of the round.
7. The round earnings of the PROPOSER.
8. The round earnings of the RESPONDER.

Selecting round for payment

Once all 30 rounds of the experiment have been completed, 1 round will be randomly chosen for payment. Each of the 30 rounds are equally likely to be chosen for payment.

Your payment for today's session will be the sum of your earnings in the round randomly chosen for payment, the initial payment of $18 and the $7 show-up fee.

Summary

- During each round, there are two possible outcomes: OUTCOME R and OUTCOME P.
- First, the PROPOSER chooses an INITIAL POSITION, which is a number between 0 and 100, inclusive.
- The PROPOSER then chooses an OFFER, which is also a number between 0 and 100, inclusive.
- The RESPONDER then sees the INITIAL POSITION and the OFFER and decides whether to accept or reject the OFFER. If the RESPONDER rejects the OFFER then the round earnings of both the RESPONDER and the PROPOSER decrease by 10. If the RESPONDER accepts the OFFER, the outcome is OUTCOME P.

  Round earnings are:

  PROPOSER: OFFER

  RESPONDER: $(-1)\cdot$ (OFFER)
- If the RESPONDER rejects the OFFER, the PROPOSER chooses a FINAL POSITION. This FINAL POSITION is any number between $0.2\cdot$ INITIAL POSITION and the INITIAL POSITION. That is, when choosing the FINAL POSITION, the PROPOSER can reduce their INITIAL POSITION by no more than 80%. The FINAL POSITION cannot be higher than the INITIAL POSITION.
- If the RESPONDER rejects the OFFER, the outcome of the round is determined using the FINAL POSITION chosen by the PROPOSER. The outcome will be OUTCOME R with probability (FINAL POSITION)%, and will be OUTCOME P with probability $(100 - \text{FINAL POSITION})\%$.

  If the outcome is OUTCOME R, round earnings are

  PROPOSER: $(-1) \cdot (\text{FINAL POSITION}) - 10$

  RESPONDER: $0 - 10 = -10$

  If the outcome is OUTCOME P, round earnings are

  PROPOSER: $(1.5) \cdot (\text{FINAL POSITION}) - 10$

  RESPONDER: $(-1) \cdot (\text{FINAL POSITION}) - 10$

Instructions for Part 2

Part 2 of the experiment is the same as Part 1, except that if the RESPONDER rejects the OFFER, the PROPOSER must choose a FINAL POSITION that is between 0.7·INITIAL POSITION and their INITIAL POSITION. That is, when choosing the FINAL POSITION, the PROPOSER can reduce their INITIAL POSITION by no more than 30%. The FINAL POSITION cannot be higher than the INITIAL POSITION.

For example, suppose the PROPOSER chooses an INITIAL POSITION of 30. Then the FINAL POSITION must be between 21 and 30.