ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN

ECONOMICS

Ciclo 36

Settore Concorsuale: 13/A1 - ECONOMIA POLITICA

Settore Scientifico Disciplinare: SECS-P/01 - ECONOMIA POLITICA

MECHANISMS FOR RESOLVING ECONOMIC DILEMMAS: EXPERIMENTAL EVIDENCE FROM BARGAINING, CONTRACTS, AND SOCIAL NORMS

Presentata da: Xin Zhang

Coordinatore Dottorato

Matteo Barigozzi

Supervisore

Maria Bigoni

Esame finale anno 2026

Mechanisms for Resolving Economic Dilemmas: Experimental Evidence from Bargaining, Contracts, and Social Norms

Xin Zhang

General Introduction

In many economic environments, individuals face dilemmas that hinder mutually beneficial outcomes due to informational frictions, incomplete contracts, or entrenched social norms. This thesis investigates how institutional mechanisms—formal or informal—can alleviate such dilemmas and improve coordination, efficiency, and welfare. Drawing from the field of experimental economics, the thesis presents three distinct but thematically connected chapters, each addressing a specific type of economic dilemma and testing mechanisms designed to resolve it.

Chapter 1 examines the classic dilemma of bargaining under incomplete information in threshold public goods provision. In such settings, mutual gains may be possible, but uncertainty about others' valuations impedes agreement. The chapter evaluates the efficacy of the Random Dictatorship (RD) mechanism, originally proposed by Myerson (1984), in resolving such coordination failures. While theory predicts that RD should facilitate agreement by selecting one party to propose a take-it-or-leave-it cost-sharing scheme, the experiment reveals that the mechanism improves preference revelation but does not significantly raise agreement rates. This outcome highlights a behavioral reluctance to contribute when not selected as dictator, even when doing so would be individually rational. Thus, while RD narrows strategic uncertainty, it does not fully overcome coordination failures in one-shot bargaining environments.

Chapter 2 turns to the domain of contract law and explores how different legal remedies influence parties' willingness and ability to renegotiate contracts after an exogenous shock. The central question is whether assigning entitlements via specific performance (a strong remedy) or damages (a weak remedy) affects the efficiency of renegotiation in the face of shocks that render the original contract suboptimal. Through a tightly controlled laboratory experiment, the chapter finds that nearly all parties attempt renegotiation and most succeed, regardless of the remedy in place. The division of the renegotiation surplus is also largely unaffected. These findings lend empirical support to the Coasean irrelevance proposition under ideal conditions—where remedies are costlessly enforceable and parties face symmetric information. However, the results also reveal bounded rationality and preference heterogeneity, suggesting that the behavioral validity of the Coase theorem may depend on contextual and institutional complexity not captured in this simplified setting.

Chapter 3 investigates how opinion leaders can influence social norm change in coordination environments characterized by pluralistic ignorance and second-order uncertainty. Using a laboratory coordination game, the chapter tests whether a strategically motivated leader—who knows the true distribution of social preferences—can help individuals reveal norm-violating preferences and shift behavior. The experiment varies the leader's ideological incentives and the social network structure (random vs. homophilous matching). Results show that when potential norm violators are in the minority, leaders with ideological incentives are more willing to endorse norm-breaking, leading to greater expression of latent preferences. Homophily amplifies this effect by reducing perceived sanction risk. These findings underscore the joint importance of leadership incentives and social structure in overcoming norm entrenchment, and provide insights into when informational influence can catalyze social change.

Together, the three chapters offer a comparative perspective on how mechanisms—ranging from institutional rules to strategic communication—can resolve economic dilemmas stemming from incomplete information, contractual rigidity, or social conformity. Each chapter employs experimental methods to isolate the effects of these mechanisms and to explore their behavioral underpinnings. The results contribute to a growing literature on mechanism design in behavioral settings, suggesting that the success of formal rules often hinges on informal incentives and cognitive constraints.

In sum, this thesis highlights that resolving economic dilemmas is not only a matter of designing optimal mechanisms in theory, but also of understanding how individuals respond to these mechanisms in practice. By combining theoretical predictions with experimental evidence, the thesis advances our understanding of how institutions can promote coordination, efficiency, and social welfare in the presence of informational and normative frictions.

Chapter 1

Random Dictatorship in Two-person Bargaining with Incomplete Information: An Experiment*

Xin Zhang

Abstract: Incomplete information often leads to failure in threshold public goods bargaining, with marked consequences for social welfare. This study examines whether the random dictatorship mechanism, proposed by Myerson (1984), may offer a remedy. In a laboratory experiment, subjects bargain under three conditions: complete information, incomplete information, and incomplete information with random dictatorship. While the mechanism improves the match between contributions and valuations, it does not raise the rate of agreement. The main reason lies in the reluctance of subjects to contribute when not selected as dictator, even when theory predicts they should. These findings suggest that, although random dictatorship can aid in information revelation, it does not overcome the coordination problem when strategic incentives are weak.

1 Introduction

Threshold public goods bargaining arises in many economic and political contexts where a project is only undertaken if participants collectively contribute enough to cover its cost. Its applications range widely, from decisions to install an elevator in an old residential building to

*The experiment has been pre-registered at AEA RCT on Feb. 28th, 2023. https://doi.org/10.1257/rct.10982-1.0. And the data collection was approved by the IRB of University of Bologna (IRB-0000990).

efforts to co-fund an intercity railway across several municipalities. Failures in such bargaining have a marked effect on social welfare. These failures are often due to incomplete information, especially about each party's private valuation of the good, which makes coordination difficult. One prominent example is climate change negotiations, where over 100 countries must share the burden of reducing carbon emissions to keep global temperature rise below 2 degrees. Even when mutual gains are possible, agreement may falter if states do not know how others value the outcome. The intricacies of such negotiations could be mitigated if each country's costs and benefits were known to all, thereby enabling a clearer and more tractable bargaining process. This study examines whether such failures can be alleviated by a mechanism proposed in the theoretical literature and tested here in a controlled laboratory setting.

To explore this question, the analysis focuses on settings with two-sided private information, where each party knows only their own valuation of the public good. In such environments, coordination is hindered not only by missing information, but by strategic uncertainty, as parties lack a common basis for agreement. As a potential remedy, the analysis considers the random dictatorship mechanism, introduced by Myerson (1984), in which one party is selected at random to dictate a cost-sharing plan, subject to the other's acceptance. In theory, such a rule may arise naturally as a focal solution when players face multiple equilibria and hold conflicting preferences—much as a coin toss may settle a stalemate between two equally insistent demands. In practice, however, such reasoning often fails to materialize, perhaps due to limited communication or bounded rationality. This study investigates whether externally enforcing this mechanism can improve outcomes when voluntary coordination breaks down. While the mechanism has been studied theoretically, its empirical performance in environments with two-sided private information remains largely unexplored.

If externally enforced random dictatorship functions as its theoretical counterpart, it serves to select one equilibrium from the many that arise in bargaining under incomplete information. Rather than requiring coordination across an infinite set of possible outcomes, the mechanism narrows the space to a single proposal determined by the selected dictator. The non-dictator, faced with a take-it-or-leave-it offer, has no incentive to reject any proposal that yields a nonnegative payoff. In equilibrium, then, any outcome acceptable to the non-dictator should be implemented without disagreement. To test this logic empirically, an online experiment was conducted with three treatments designed to isolate the effects of incomplete information and of externally imposed random dictatorship. In the complete-information treatment ("Complete"), each participant knows both valuations. In the incomplete-information treatment ("Incomplete"), each knows only their own valuation. The random dictatorship treatment ("Dictator") builds on the incomplete-information setting, but introduces an externally imposed mechanism: one participant is randomly selected as the dictator and proposes a cost-sharing plan, which the other may accept or reject. To isolate the impact of enforced random dictatorship, bargaining is conducted in a one-shot simultaneous setting, which prevents the mechanism from arising naturally through communication or coordination.

The experimental results confirm that incomplete information reduces agreement in threshold public goods bargaining, as predicted. Throughout the paper, "agreement" refers to situations in which the two parties successfully coordinate to implement the joint project — that is, when their combined contributions meet or exceed the threshold required to fund it. However, externally imposing random dictatorship does not significantly increase the overall agreement rate. While the mechanism improves the alignment between willingness to contribute and private valuations—suggesting that it helps elicit individual preferences more truthfully—it falls short of restoring coordination. The main reason appears to be a reluctance to contribute when not selected as dictator, even when doing so would be individually rational. Many participants deviated from equilibrium play by making overly generous offers or rejecting proposals that would have left them no worse off. This reluctance to follow the strategic logic of the game appears to limit the mechanism's effectiveness in restoring agreement.

The remainder of the paper is structured as follows. Section 2 reviews the related literature. Section 3 presents the experimental design and theoretical predictions. Section 4 reports the results. Section 5 discusses their implications and concludes.

2 Literature Review

Foundational work on bargaining under complete information (Nash, 1950; Kalai and Smorodinsky, 1975) has been extended to environments with incomplete information, where equilibrium multiplicity becomes a central challenge. Myerson (1984) develops a general two-person bargaining framework with private information and introduces Random Dictatorship as a coordination axiom to select among multiple incentive-efficient, incentive-compatible, and individually rational Bayesian equilibria. Under Random Dictatorship with equal bargaining power, a chance move selects one party; the implemented outcome is that party's most-preferred equilibrium in the feasible set. In a threshold public good environment, this leads to a screening solution: the selected party proposes a cost-sharing plan, which the other accepts if and only if the implied payment does not exceed her privately known value.

While subsequent theoretical work has acknowledged the equilibrium multiplicity emphasized by Myerson (1984), and occasionally invoked the Random Dictatorship axiom as a normative selection principle (e.g., Herrero and Martinelli, 1998; Chatterjee and Dutta, 1993), this approach has not led to a unified strand of models or sustained empirical testing. In the broader literature, attention has largely shifted toward equilibrium refinements (e.g., Cho and Kreps, 1987), dynamic bargaining under asymmetric information (e.g., Ausubel, Cramton, and Deneckere, 2002), and models of signaling or learning in bargaining environments (e.g., Fudenberg and Villas-Boas, 2007). These lines of research have primarily addressed bilateral trade or wage bargaining settings, and have typically not engaged with public good bargaining or axiomatic selection mechanisms like Random Dictatorship.

In this context, the Random Dictatorship axiom remains a tractable and incentive-compatible benchmark for bargaining with incomplete information. Its predictions are sharp and its structure allows for clear implementation in environments characterized by coordination failure, such as threshold public goods. Yet despite its theoretical appeal, Random Dictatorship has received little empirical attention and remains largely untested in controlled environments. This absence of systematic evaluation—rather than any contradiction or refutation—suggests that Myerson's (1984) solution concept remains theoretically salient and warrants further investigation through experimental methods. This study builds on Myerson's theoretical framework by testing Random Dictatorship in a novel experimental environment involving two-sided private values and collective provision.

Bargaining under incomplete information, by contrast, has attracted sustained experimental attention. A prominent strand of this literature builds on the framework of Chatterjee and Samuelson (1983), which derives a unique Bayesian equilibrium in a bilateral trade setting under the constraint of linear strategies. Radner and Schotter (1989) find that experimental outcomes in sealed-bid bargaining environments are broadly consistent with this linear equilibrium. Several other studies explore how information asymmetries influence bargaining dynamics. Rapoport et al. (1998) show that traders can strategically exploit informational advantages to increase their surplus. Valley et al. (2002) find that both face-to-face communication and written message exchange can mitigate the inefficiencies caused by private information. Building on this, Ellingsen et al. (2009) demonstrate that communication only improves coordination when players have previously cooperated in a pre-play Prisoner's Dilemma, and Lundquist et al. (2009) find that individuals exhibit an aversion to lying in pre-bargaining cheap talk. Complementary evidence from real-world settings comes from studies such as Backus et al. (2020), who analyze eBay negotiation data, and Larsen (2021), who studies used car auction markets.

Beyond the domain of bilateral trade, several experiments have explored other bargaining environments with private information — including settings more directly related to the current study, such as public good provision and mechanism-based bargaining. Mitzkewitz and Nagel (1993) study ultimatum games with incomplete information; Hoffmann et al. (2015) investigate the effects of pre-play commitments; and Camerer et al. (2019) analyze unstructured bargaining with private valuations using machine learning. Among these, the study closest to the current design is Forsythe et al. (1991), who experimentally test the Random Dictatorship axiom in a divide-the-dollar game with one-sided incomplete information about the size of the pie, in contrast to the two-sided private-value setting studied here. Their findings show that disagreement becomes more likely when the pie is smaller, consistent with Myerson's predictions, and that implementing Random Dictatorship reduces inefficiency, though the selected dictator tends to appropriate a larger share of the surplus.

This study extends that line of inquiry by applying the random dictatorship mechanism to settings with two-sided incomplete information, and by embedding it in a threshold public goods environment following Myerson (1984).

The logic behind random dictatorship — where one party's proposal may be selected over another's—bears a structural resemblance to mechanisms studied in final-offer arbitration (FOA). FOA is a well-studied procedure in law and economics, particularly within the context of labor disputes. Under FOA, each party submits a final proposal, and a third-party arbitrator must choose one of the two offers in its entirety, without modification. This structure is intended to discourage extreme positions and incentivize convergence toward fair or moderate outcomes (Brams and Merrill, 1983). Unlike Random Dictatorship, which relies on random selection and allows the unchosen party to reject the proposal, FOA involves an external evaluator and results in a binding decision, regardless of either party's willingness to accept. Moreover, FOA is typically applied in adversarial, zero-sum disputes (e.g., wage bargaining or contract resolution), whereas Random Dictatorship is implemented here in a cooperative setting involving the voluntary provision of a threshold public good. These differences in selection rule, enforceability, and context distinguish the two mechanisms, even though both involve the choice of one party's offer over another's.

Recent contributions in the FOA literature have expanded its theoretical and empirical scope. Farber (1980) models FOA with risk-averse agents and highlights how enforcement and asymmetric information affect proposal strategies. More recently, experimental studies have examined how the design of arbitration procedures shapes outcomes: Deck and Farmer (2007) show that whole-package FOA leads to more extreme offers and greater inefficiency relative to issue-by-issue mechanisms. Dickinson (2009) finds that beliefs about the opposing party's strategy significantly affect outcomes, emphasizing the importance of information structure. A broader overview by Dickinson (2020) highlights growing interest in applying behavioral and experimental methods to arbitration design. This study contributes to the FOA literature by introducing a related but distinct mechanism—random dictatorship—into a cooperative bargaining environment with two-sided private information. While not formally an FOA setting, the strategic structure invites comparison and offers new insights into how selection-based mechanisms perform under asymmetric information when acceptance is voluntary. In doing so, this study contributes both to the experimental validation of random dictatorship and to the design-oriented literature on mechanisms for resolving bargaining under private information.

3 Experiments

3.1 Experimental Setup

The experiment implements a one-shot threshold public goods bargaining game with two-sided private information, adapted from the example in Myerson (1984, Section 10)[1]. Two players must decide whether to invest in a joint project that generates value for both, but only if they collectively contribute at least 100 tokens—the cost of the project.

Each player's benefit from the project depends on their privately known type: High or Low. A High-type player values the project at 90 tokens; a Low-type player values it at 30. Each participant is informed that both they and their match are independently assigned types, with a 50% probability of being High and a 50% probability of being Low. This probability distribution is common knowledge.

The experiment includes three treatments: a Complete Information treatment ("Complete"), in which players observe each other's valuations; an Incomplete Information treatment ("Incomplete"), where each player knows only their own valuation; and a Random Dictatorship treatment ("Dictator"), which retains private information but assigns one player as proposer

[1] This experimental design adapts the example from Myerson (1984, Section 10) in three key respects. First, while the original model features one-sided private information (only player 1 has uncertain valuation), the current design introduces two-sided private information, where both players' valuations are private. Second, the original model assigns a 90% probability to the high type and 10% to the low type for player 1; in contrast, the present design uses a symmetric 50–50 distribution for both players to simplify the experimental structure. Third, several contextual elements have been modified to maintain a neutral framing: for instance, "the road" in Myerson's example is replaced with "a joint project."

with the power to make a take-it-or-leave-it offer. These treatments are designed to isolate the effects of information asymmetry and to evaluate whether externally enforcing the random dictatorship mechanism mitigates coordination failure:

- Comparing "Complete" and "Incomplete" identifies whether and how incomplete information reduces agreement rates and efficiency.
- Comparing "Incomplete" and "Dictator" isolates the effect of externally imposed random dictatorship as a potential remedy for the coordination failures induced by incomplete information.

All treatments are implemented in a one-shot, anonymous matching environment to eliminate repeated-game effects or strategic learning.

Next, I introduce in detail the strategy space of participants across the three treatments. A summary of the decision steps is presented in Table 1.

In the "Complete" treatment, each participant observes both their own and their partner's type before making a decision. In the "Incomplete" treatment, each participant knows only their own type and is informed that their match may be either type with equal probability. In both treatments, participants simultaneously propose how much they are willing to contribute to the joint project. The proposal must be an integer ranging from 0 to the participant's own valuation: that is, 0–90 for High-type participants and 0–30 for Low-type participants.

If the total contribution from both players is at least 100 tokens, the project is implemented; such cases are referred to as "agreements." One of the two proposals is then randomly selected and implemented: the proposer pays the amount they proposed, and the matched player pays the residual cost. If the total is below 100, the project is not implemented and both players earn zero.

The "Dictator" treatment retains the same private information structure as "Incomplete" but modifies the bargaining process. One player is randomly selected as the proposer ("dictator") and submits a proposed contribution, again choosing an integer between 0 and their own valuation. The matched player then decides whether to accept the proposal by covering the remaining cost.

To implement this structure using the strategy method, all participants make both proposer and responder decisions in advance. First, each participant states their proposed contribution if selected as the proposer. Then, to elicit their willingness to accept if not selected, each participant completes a staircase procedure consisting of binary accept/reject choices. This sequence of hypothetical offers indicates how much the match would contribute, and thus how much the participant would need to pay - with the required payment ranging from a minimum of 10 tokens up to the participant's own valuation (i.e., 90 for High-types and 30 for Low-types). The process identifies the participant's maximum acceptable contribution (MAC).

If the proposer's stated contribution and the responder's MAC sum to at least 100 tokens, the project is implemented. As in the other treatments, I refer to such outcomes as "agreements." The full logic of the staircase procedure is described in Appendix C.[2]

Table 1: Summary of Decision Steps Across Treatments

Treatment	**Information Provided**	**Proposal Decision**	**Acceptance Decision (MAC)**
"Complete"	Own value and match's value	Yes	No
"Incomplete"	Own value only	Yes	No
"Dictator"	Own value only	Yes	Yes

Note: In the "Dictator" treatment, the acceptance decision is elicited using a staircase procedure to determine the maximum acceptable contribution (MAC). See Section 3.1 and Appendix C for details.

[2] The staircase procedure works as follows (see Appendix A.2 for an illustrative example). Participants are first asked whether they would accept a proposal in which their match contributes a specific amount toward the joint project, leaving them to cover the remainder. If the participant selects "Accept," the next question reduces the match's contribution, increasing the amount the participant would need to pay. If they select "Reject," the next question increases the match's contribution, reducing the participant's share. This iterative process continues until the participant's maximum acceptable contribution is identified. In the figures shown in Appendix C, participant responses are labeled as "A" (Accept) and "R" (Reject) for brevity.

3.2 Procedures

The experiment was conducted using Qualtrics and involved 320 participants recruited from the subject pool of the BLESS lab at the University of Bologna. The experiment was conducted in English. A between-subject design was used: each participant was randomly assigned to one of three treatments—Complete Information ("Complete"), Incomplete Information ("Incomplete"), or Random Dictatorship ("Dictator")—and was not informed about the existence of the other treatments.

Participants could access the experiment remotely at any time on the day it was launched. After reading an information sheet and providing informed consent, they received neutrally framed instructions and completed comprehension questions. They could proceed only after answering all questions correctly. The full experimental instructions are included in Appendix A.

Each participant was informed of their own value (either High or Low) and asked to make a proposal indicating how much they were willing to contribute to the joint project. In the "Dictator" treatment, participants also completed a sequence of accept/reject decisions, implemented via the staircase procedure (see Section 3.1 and Appendix C), to elicit their maximum acceptable contribution. The strategy method was used: all decisions were made in advance, without interaction. This approach was appropriate given the one-shot nature of the experiment, which ruled out repeated-game dynamics and reputation effects, and it allowed for the collection of richer data. In particular, the strategy method enabled the measurement of each participant's full acceptance profile in "Dictator", rather than a single binary response. The structure of type assignment across treatments is summarized in Table 2.

After completing their decision tasks, all participants filled out a short questionnaire (reproduced in Appendix B) collecting demographic information and measures of individual characteristics, including risk preference, fairness preference, and altruism.

Only after all participants had completed the experiment (including the questionnaire), they were manually matched by the experimenter using pre-specified random seeds[3]. For each matched pair, the agreement outcome and implemented cost-sharing plan (if any) were determined from the submitted decisions. If the two proposals (in "Complete" and "Incomplete") were together sufficient to cover the project cost, one of them was randomly selected and implemented as the final cost-sharing plan. In "Dictator", if the selected dictator's proposal and the non-dictator's maximum acceptable contribution together met or exceeded the project cost, the dictator's proposal was implemented. If no agreement was reached, the project was not implemented and both participants received zero. The results were communicated to participants via a pre-coded automated email on Qualtrics.

Payments were calculated in tokens and converted to euros at a rate of 5 tokens = €1. The final payment included the payoff from the implemented decision plus a €1 show-up fee, and was transferred via PayPal. The average payment was €4.1, with a minimum of €1 and a maximum of €17. Participants were informed in advance that they would receive their payment within one week, and all payments were processed accordingly. The experiment was approved by the Institutional Review Board of the University of Bologna (IRB-0000990) and preregistered with the AEA RCT Registry (ID: https://doi.org/10.1257/rct.10982-1.0). The full materials and decision screens are available in the online appendix.

3.3 Theoretical Predictions

This section presents theoretical benchmarks for each treatment. These predictions serve as a baseline to evaluate whether and how the information structure and bargaining mechanisms influence agreement rates and efficiency. For the Complete Information and Random Dictatorship treatments, theoretical predictions are derived from well-established bargaining

[3]In "Complete", there are 26 pairs of (High, High) and 26 pairs of (High, Low). In "Incomplete", there are 13 pairs of (High, High), 26 pairs of (High, Low), and 13 pairs of (Low, Low). In "Dictator", there are 14 pairs of (High, High), 28 pairs of (High, Low), and 14 pairs of (Low, Low).

[4]The number of subjects in "Dictator" was intended to match those in "Complete" and "Incomplete" (104 in total). Based on past experience, 110 participants were assigned to both the High and Low types, anticipating some attrition. However, all High-type subjects completed the survey, while some Low-type subjects did not. As a result, the final number of participants in "Dictator" was 112.

Table 2: The Assignment of Subjects for Each Treatment

Treatment	Your type	Your match's type	# of subjects
"Complete"	High	High	52
	High	Low	26
	Low	High	26
"Incomplete"	High	Either High or Low, 50% - 50% probability	52
	Low	Either High or Low, 50% - 50% probability	52
"Dictator"	High	Either High or Low, 50% - 50% probability	56[4]
	Low	Either High or Low, 50% - 50% probability	56

Note: In "Complete", participants were informed of both their own and their match's type, and no (Low, Low) pairs were assigned. In "Incomplete" and "Dictator", participants were informed only of their own type, and were told that their match's type would be randomly assigned with a 50% chance of being High or Low.

models: the Nash bargaining solution and the equilibrium selection criteria developed in Myerson (1984), respectively. For the Incomplete Information treatment, however, no unique equilibrium is predicted due to the strategic uncertainty introduced by private information and the absence of a coordination mechanism. Instead, this treatment will be discussed in terms of comparative expectations relative to the other two settings.

Complete Information ("Complete")

Under complete information, I refer to the Nash bargaining solution (Nash, 1950), which maximizes the product of the players' net payoffs:

$$\pi = (u_1 - c_1) \cdot (u_2 - c_2),$$

where u_i is player i's benefit from the project (i.e., their type value), and c_i is their contribution. This solution equalizes the net payoff between players whenever the project is implemented. Except when two Low-type players are matched (each valuing the project at 30), the project should be implemented and the surplus shared efficiently.

The resulting theoretical predictions for "Complete" are summarized in Table 3. While Table 3 includes the (30, 30) case for completeness, this match type was excluded in the "Complete" treatment, since the project cannot be implemented when both players are Low-type.

Table 3: Theoretical Predictions for Complete Information ("Complete")

Your value, Match's value	Project Realization	Contributions	Payoffs
$90, 90$	Yes	$50, 50$	$40, 40$
$90, 30$	Yes	$80, 20$	$10, 10$
$30, 90$	Yes	$20, 80$	$10, 10$
$30, 30$	No	-	$0, 0$

Random Dictatorship ("Dictator")

In the "Dictator" treatment, one participant is randomly selected to be the "dictator", whose proposed cost-sharing plan is implemented if and only if the match accepts it. From a theoretical standpoint, the "Dictator" mechanism creates an asymmetric bargaining environment under incomplete information. Each participant submits a proposal assuming they may be selected as dictator, and simultaneously provides a set of binary accept/reject decisions (via the staircase method) assuming they are not selected.

Following Myerson (1984), the set of equilibria in two-person bargaining with private values is generally large, even under incentive efficiency and individual rationality constraints. Among these, the strategy profile selected in the "Dictator" treatment corresponds to the equilibrium that maximizes each player's expected payoff. Specifically, proposers offer the minimum amount they believe will be accepted, while responders accept any proposal up to the value they assign to the joint project. In our implementation, where values are either 90 or 30 and contributions are restricted to integers, a High-type proposer (value=90) offers 11 tokens—the smallest integer strictly more than 10—since it guarantees acceptance from another High-type player (who accepts to contribute up to 89). Low-type responders (value=30) accept to contribute up to 29.

Based on this logic, the predicted outcomes are:

- If both players are High-type, the project is always implemented.
- If the match is High–Low or Low–High, the project is implemented only when the Low-type is the proposer (i.e., selected as dictator).
- If both players are Low-type, the project is never implemented.

Since the dictator is randomly selected, the expected implementation rate and payoff in mixed-type pairs (High–Low or Low–High) is 50%. The predicted payoffs, assuming that all players follow the equilibrium strategies described above, are summarized in Table 4.

Table 4: Theoretical predictions under Random Dictatorship ("Dictator")

Your value	Your proposal	Your maximum acceptable contribution	Your match's value	Project Realization	Your expected Payoff
90	11	89	90	100%	40
			30	50%	0.5
30	11	29	90	50%	9.5
			30	0%	0

Note: In theory, a player not selected as proposer should accept any contribution strictly less than their valuation (i.e., < 90). Since contributions are restricted to integers in the experiment, the maximum acceptable contribution for a High-type becomes 89. Consequently, proposers optimally offer 11 tokens rather than 10.

Incomplete Information ("Incomplete")

The Incomplete Information treatment preserves the core structure of the bargaining game used in the Complete Information condition, but removes common knowledge of types. Each participant knows their own value but not their match's type, and submits a proposal without any feedback or communication. This setting introduces strategic uncertainty on both sides and eliminates the external mechanism used in "Dictator" to coordinate on a single equilibrium.

Theoretically, this environment admits a wide range of Bayesian equilibria, and no unique prediction can be derived without additional assumptions. Coordination is especially difficult in mixed-type matches (High–Low or Low–High), where players may form inconsistent expectations about their partner's contribution or willingness to pay. Even in same-type matches, equilibrium play may break down if participants believe their match might behave cautiously due to uncertainty.

Compared to "Dictator", the "Incomplete" treatment lacks the equilibrium selection force of the dictatorship rule. While some players may attempt to mimic efficient strategies (e.g., High-types proposing around 50), without assurance of their match's type or a dominant role, such strategies are risky. This uncertainty is expected to reduce the agreement rate and lower overall efficiency compared to both "Complete" and "Dictator".

3.4 Hypotheses

Based on the theoretical benchmarks presented in Section 3.3, the following hypotheses are formulated to guide the empirical analysis. These hypotheses focus on two main outcome dimensions: (i) agreement rates across treatments and types, and (ii) willingness to contribute, measured by proposed contributions (PC) and maximum acceptable contributions (MAC).

Agreement Rate Hypotheses

Hypothesis 1. *Almost all bargaining pairs in the Complete Information ("Complete") treatment reach agreement.*

Hypothesis 2. *The agreement rate in the Random Dictatorship ("Dictator") treatment is lower than in "Complete".*

- *For (High, High) pairs: agreement rate in "Dictator" ≈ "Complete".*

- *For (High, Low) pairs: agreement rate in "Dictator" < "Complete".*

Hypothesis 3. *The agreement rate in the Incomplete Information ("Incomplete") treatment is lower than in "Dictator".*

- *For (High, High) pairs: agreement rate in "Incomplete" < "Dictator".*
- *For (High, Low) pairs: agreement rate in "Incomplete" < "Dictator".*

Willingness to Contribute Hypotheses

Let $\text{PC}_t(s, q)$ denote the proposed contribution in treatment $t \in \{\text{"Complete"}, \text{"Incomplete"}, \text{"Dictator"}\}$ made by a subject of type $s \in \{\text{High}, \text{Low}\}$ under the belief or information that the match's type is $q \in \{\text{High}, \text{Low}, \text{Unknown}\}$. That is, "Unknown" refers to the cases ("Incomplete" and "Dictator") where the subject does not observe the match's type. Let MAC_s denote the maximum acceptable contribution by a subject of type s in "Dictator". Then:

Hypothesis 4. *Comparison of willingness to contribute across treatments:*

- *For High-type players:* $MAC_H > PC_{Complete}(High, Low) > PC_{Incomplete}(High, Unknown) > PC_{Complete}(High, High) > PC_{Dictator}(High, Unknown)$
- *For Low-type players:* $MAC_L > PC_{Complete}(Low, High) \approx PC_{Incomplete}(Low, Unknown) > PC_{Dictator}(Low, Unknown)$

These comparisons reflect the theoretical predictions developed in Section 3.3. While the proposed contribution (PC) values in the "Complete" and "Dictator" treatments are derived from model-based equilibria, the "Incomplete" treatment lacks a unique theoretical prediction due to strategic uncertainty and the absence of a coordination mechanism. For this reason, the inequality signs involving the "Incomplete" should not be interpreted as equilibrium predictions, but rather as comparative expectations grounded in players' incentives and the structure of information.

In particular, High-type players in "Incomplete" must consider the possibility of being matched with either a High- or a Low-type partner. As a result, their proposals are expected to fall between the two "Complete" benchmarks: higher than in "Complete" when matched with a High-type, but lower than in “Complete” when matched with a Low-type.

Similarly, although Low-type players in “Incomplete” do not observe their match’s type, they know that agreement is only feasible if their match is a High-type. Accordingly, they may condition their proposal as if the match were a High-type. This suggests that their willingness to contribute should resemble that of Low-type players in “Complete” when matched with a known High-type. However, the lack of common knowledge and strategic certainty may still lead to behavioral deviations from that benchmark.

This reasoning aligns with standard practice in experimental economics when theoretical predictions are ambiguous or indeterminate. The inclusion of the “Incomplete” treatment in Hypothesis 4 therefore reflects structured behavioral expectations rather than formal equilibrium outcomes.

Hypothesis 5. *Decision logic for contribution decisions:*

- *In “Complete”, proposed contributions depend on both the subject’s and the match’s type.*
- *In “Dictator”, proposers are expected to propose 11 tokens regardless of type, consistent with the payoff-maximizing equilibrium, while responders’ MAC should correspond closely to their private value.*
- *In “Incomplete”, proposed contributions depend only on the subject’s own type.*

4 Results

This section presents the main experimental results. It begins by comparing agreement rates and willingness to contribute across the three treatments for each type match. This descriptive analysis provides a first test of the theoretical predictions outlined in Section 3.3 and the hypotheses formulated in Section 3.4. It then turns to regression analysis to examine how subjects’ decisions relate to their assigned types and to the institutional features of each treatment. Summary statistics on participants’ demographics and social preferences are reported in Appendix D.

4.1 Agreement Rate

To calculate agreement rates in each treatment, subjects were randomly paired using a pre-specified seed number 2023[5] in R.

In the "Complete" treatment, type assignments pre-determined match types: 26 (High, High) pairs and 26 (High, Low) pairs were formed, with no (Low, Low) matches assigned, as such pairs cannot reach agreement.

In the "Incomplete" treatment, all participants made decisions without knowing their match's type. To allow estimation of agreement rates by match type, pairings were generated ex post using random matching and revealed types. This resulted in 26 (High, High) pairs, 52 (High, Low) pairs, and 26 (Low, Low) pairs. Since (Low, Low) matches cannot result in agreement, they are excluded from the analysis.

This procedure implies that each of the 52 High-type participants was matched twice (once in a (High, High) pair and once in a (High, Low) pair), while each Low-type participant was matched once. However, this duplication occurred only during the analysis stage, was not visible to participants, and does not affect incentives or behavior.

Crucially, participants were paid for only one randomly selected match, as stated in the instructions. Therefore, the procedure does not generate income effects or strategic distortion. Since all decisions were made simultaneously using the strategy method, and without feedback or interaction, repeated use of High-type decisions does not compromise statistical independence.

Inference is based on comparisons within each match type across treatments (e.g., comparing (High, Low) matches in "Complete", "Incomplete", and "Dictator"). No comparisons are made between different match types within a treatment. For this reason, the duplication of High-type observations does not bias the results. The same logic applies to the Dictator treatment, where ex post pairing produced 28 (High, High) and 56 (High, Low) matches, and

[5] The use of random matching with seed 2023 in R was pre-specified in the pre-analysis plan.

all (Low, Low) pairs were again excluded from analysis.

Figure 1 displays the average agreement rates across treatments by match type. The y-axis shows agreement rates (from 0 to 1), and the x-axis presents combinations of treatments and type matches. Each bar shows either the experimentally observed (blue) or theoretically predicted (light blue) agreement rate for a given type pair in a specific treatment. The green lines indicate 95% confidence intervals.

In the "Complete" treatment, the overall agreement rate is slightly but significantly lower than 80%[6] ($p = 0.04$, one-sided Binomial test, $n = 52$). Among (High, High) pairs specifically, the agreement rate is not significantly different from 80% ($p = 1.00$, two-sided Binomial test, $n = 26$). For this match type, agreement rates do not differ significantly across treatments ($p > 0.7$, Chi-squared test; "Complete": $n = 26$, "Incomplete": $n = 26$, "Dictator": $n = 28$).

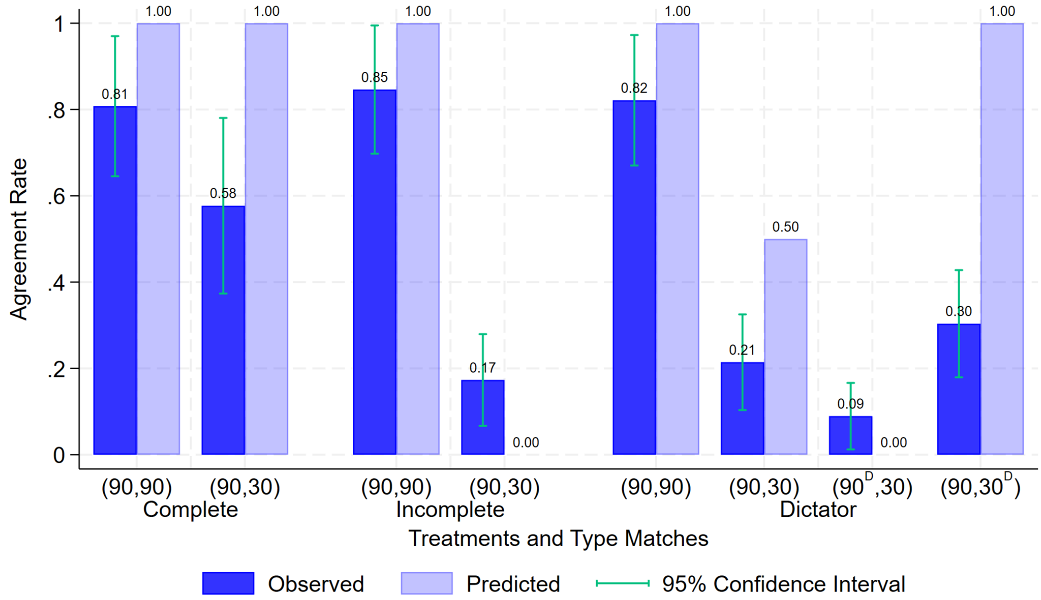
In contrast, for (High, Low) pairs, the agreement rate in "Complete" is significantly higher than in both "Dictator" and "Incomplete" ($p < 0.01$ for both comparisons, Chi-squared test; "Complete": $n = 26$, "Incomplete": $n = 52$, "Dictator": $n = 56$). There is no significant difference between "Dictator" and "Incomplete" in this match type ($p = 0.59$, Chi-squared test). These findings suggest that private information about the match's valuation substantially reduces agreement in heterogeneous-type pairs. Moreover, the enforcement of random dictatorship does not significantly improve agreement relative to incomplete information without such a mechanism.

4.2 Willingness to Contribute

This section reports how the subjects' willingness to contribute varies across treatments. In "Complete" and "Incomplete", this variable is measured by the proposal made by each subject. In "Dictator", willingness to contribute is elicited for both the proposer and the responder: the proposer makes a proposal, and the responder indicates the maximum acceptable contribution

[6] Here the agreement rate is tested against 80% rather than 100%, as 80% was set as the benchmark threshold for near-universal agreement in the pre-analysis plan.

Figure 1: Agreement rates across treatments by match type



Note: For the "Complete" and "Incomplete" treatments, each includes four bars: observed and predicted agreement rates for (High, High) matches, labeled as $(90, 90)$, and (High, Low) matches, labeled as $(90, 30)$. In the "Dictator" treatment, eight bars are shown to reflect the role of the proposer. In addition to the general (High, Low) match type, the bars are split into $(90^D, 30)$, where the High-type participant was selected as dictator, and $(90, 30^D)$, where the Low-type was selected as dictator.

(MAC) using the staircase procedure. Figure 2 displays the average values of willingness to contribute across treatments, by participant type and match type. Willingness to contribute is measured as:

- Proposed contributions (blue bars) in all treatments, and
- Maximum acceptable contributions (MAC) (orange bars), elicited only in the "Dictator" treatment.

As in Figure 1, both experimental data (solid blue/orange) and theoretical predictions (light blue/light orange) are shown.The x-axis presents type-match categories, and the y-axis shows average contribution values. Black lines indicate 95% confidence intervals.

The following empirical ranking is derived from the data[7]:

$$MAC_H \approx PC_{Complete}(\text{High}, \text{Low}) > PC_{Incomplete}(\text{High}, \text{Unknown}) > PC_{Complete}(\text{High}, \text{High})$$

[7] Sample sizes for the Mann–Whitney U tests are as follows: $MAC_H : n = 56$; $PC_{Complete}(High, Low)$:

$\approx PC_{Dictator}(\text{High}, \text{Unknown})$, and

$MAC_L > PC_{Complete}(Low, High) \approx PC_{Incomplete}(Low, Unknown) > PC_{Dictator}(Low, Unknown)$.

Here, all ">" relationships are supported by Mann–Whitney U tests with $p < 0.01$, and all "≈" relationships correspond to tests with $p > 0.20$. These comparisons provide a structured empirical overview of how information structure and role assignment influence willingness to contribute.

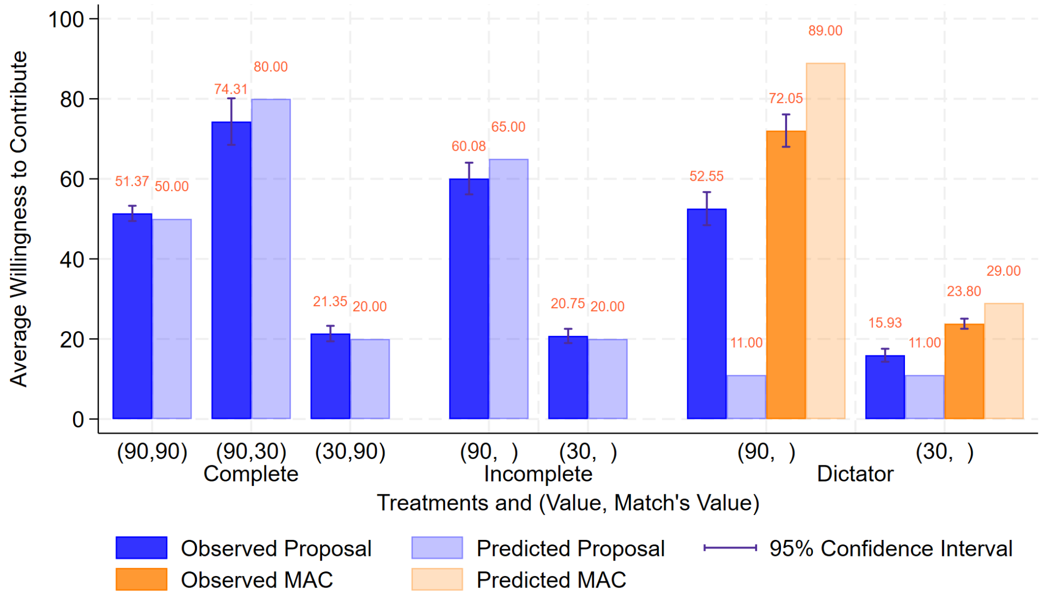
The ranking of Low-type players is the same as predicted by the theory, while for the ranking of High-type players, there are two differences from the theoretical predictions. Firstly, MAC_H is not significantly higher than $PC_{Complete}(High, Low)$, which means that although "Dictator" can increase High-type subjects' willingness to contribute, the maximum they are willing to accept contributing is no more than the amount they would propose when they know their match is Low-type. The second difference is that $PC_{Complete}(High, High)$ is not significantly higher than $PC_{Dictator}(High, Unknown)$, meaning that High-type subjects did not adopt the theoretical strategy of proposing 10; instead, they behaved as if they knew their match was High-type.

Moreover, in "Dictator", High-type players' maximum acceptable contribution (average: 72.1, theory: 89) is much lower than the predicted level, while their proposed contribution as proposers (average: 52.6, theory: 10) is much higher. A similar pattern holds for Low-type players in "Dictator": average proposal 15.9 > 10 (theory), and average MAC 23.8 < 29 (theory).

Although there is a significant discrepancy between the observed and predicted willingness to contribute in the Random Dictatorship treatment—both in terms of proposals and maximum acceptable contributions (MACs)—these deviations may reflect participants' attempts to best respond to their expectations about the match's behavior. As shown in Figure 3, while most observed choices by High-type participants (red dots) deviate substantially from

$n = 26$; $PC_{Incomplete}(High, Unknown) : n = 52$; $PC_{Complete}(High, High) : n = 26$; $PC_{Dictator}(High, Unknown) : n = 56$; $MAC_L : n = 56$; $PC_{Complete}(Low, High) : n = 26$; $PC_{Incomplete}(Low, Unknown) : n = 52$; $PC_{Dictator}(Low, Unknown) : n = 56$.

Figure 2: Comparison of Willingness to Contribute



Note: Each bar represents the average willingness to contribute for a given type match. Type matches are labeled as (Participant's Value, Their Match's Value known by the Participant). In the Complete treatment, where types are known, there are three match types:

- $(90, 90)$: High-type participants matched with High-type
- $(90, 30)$: High-type participants matched with Low-type
- $(30, 90)$: Low-type participants matched with High-type

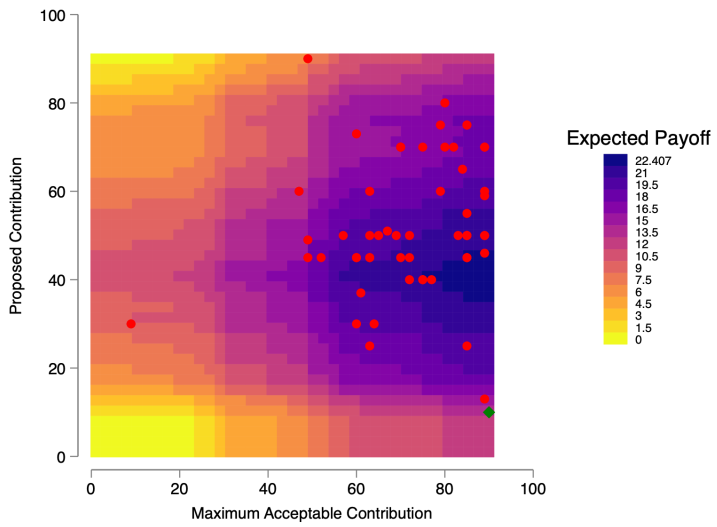
In the Incomplete treatment, match types are not known, so labels indicate only the participant's value:

- $(90, \ )$: High-type with unknown match
- $(30, \ )$: Low-type with unknown match

In the Dictator treatment, participants face the same uncertainty as in Incomplete. The same type labels are used for proposed contributions (blue), while additional orange bars show elicited maximum acceptable contributions (MAC).

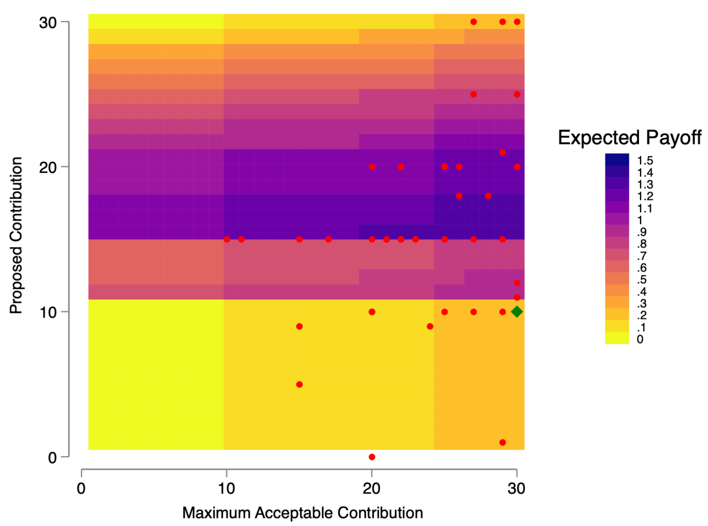
the theoretical prediction (green diamond), they are located in regions corresponding to high expected payoff. A similar pattern is observed in Figure 4: more than half of the Low-type participants' decisions fall in areas where the match's expected contribution would allow for payoff higher than the one they would obtain by abiding the theoretical prediction.

Figure 3: Response Patterns of High-type Participants in Random Dictatorship



Note: Each red dot represents a High-type participant's combination of (Proposal, MAC) in "Dictator". The green diamond denotes the theoretical prediction. The background shading reflects expected payoff levels, with darker areas indicating higher expected payoffs.

Figure 4: Low Type's Real Choice and Expected Payoff in "Dictator"



Note: Each red dot represents a Low-type participant's combination of (Proposal, MAC) in "Dictator". The green diamond denotes the theoretical prediction. The background shading reflects expected payoff levels, with darker areas indicating higher expected payoffs.

However, a subset of Low-type participants (9 out of 56)—those who submitted both low proposals and low MACs—are situated neither near the theoretical prediction nor in regions associated with favorable match behavior. One natural hypothesis is that such decisions may reflect risk aversion or inequality aversion, both of which could lead participants to forgo potentially profitable agreements in order to avoid unequal or uncertain outcomes. To examine this possibility, the next subsection incorporates questionnaire-based measures of risk preferences and distributive preferences into a regression analysis. As shown below, however, these factors provide limited explanatory power, suggesting that standard social-preference motives alone cannot account for the observed deviations from equilibrium play.

4.3 Regression Analysis on Willingness to Contribute

This subsection further investigates the determinants of subjects' willingness to contribute through individual-level regression analysis.

Table 5 presents the OLS regression results based on the following model specification:

$$WTC_i = \alpha + \beta_1\, value_i + \beta_2\, value\_m_i + \sigma D_i + \mu S_i + \epsilon_i,$$

where WTC_i denotes subject i's willingness to contribute (measured as either *Proposal* or *MAC*), $value_i$ is subject i's own value from the joint project, and $value\_m_i$ is the value of their matched participant. In the Incomplete and Dictator treatments, participants did not observe their match's type when making decisions, and matching was conducted only after all choices were submitted. Therefore, the partner's valuation ($value_m$) was not known or determined at the time of decision-making, and is excluded from the regression specification for these treatments. The vectors D_i and S_i include, respectively, the subject's demographic controls and responses to social preference measures. In Table 5, the variables `risk_self`, `risk_bomb`, `fair1`, `fair2`, `donation`, and `share_self` correspond to Questions 5–10 in the post-experimental questionnaire (Appendix B).

The estimated coefficients on *value* and *value_m* indicate that in the Complete Information treatment, proposed contributions increase with a subject's own value and decrease with the match's value. This pattern aligns with strategic reasoning: when a subject expects their match to benefit more from the joint project, they reduce their own contribution, anticipating the match will contribute more.

In the "Incomplete" and "Dictator" treatments, only a participant's own value is observed, and accordingly, only *value* has a significant positive effect on *Proposal*. This contrasts with the theoretical prediction under random dictatorship, where proposals should be constant across values. The observed variation suggests that participants were adjusting to perceived uncertainty in their match's behavior, rather than strictly following the equilibrium strategy.

Furthermore, in "Dictator", the coefficient on *value* is even larger when MAC is the dependent variable, implying that MAC responses more directly reflect participants' private valuations than their proposals, which may be distorted by strategic concerns. Notably, the constant term is significant only in the "Complete", suggesting that subjects may anchor their decisions on a reference point and adjust their contributions based on both their own and their match's value. By contrast, in the "Incomplete" and "Dictator" treatments, willingness to contribute appears to scale proportionally with one's own value.

Table 5: The Regression of Willingness to Contribute[8]

Treatment	Complete			Incomplete			Dictator					
	Dependent Variable: Proposal									Dependent Variable: MAC		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
value	0.50^{***}	0.50^{***}	0.51^{***}	0.66^{***}	0.64^{***}	0.65^{***}	0.61^{***}	0.61^{***}	0.61^{***}	0.80^{***}	0.81^{***}	0.80^{***}
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)
value_m	-0.38^{***}	-0.39^{***}	-0.40^{***}									
	(0.04)	(0.04)	(0.04)									
age	no	yes	yes	no	yes	yes	no	yes	yes	no	yes	yes
gender	no	yes	yes	no	yes	yes	no	yes	yes	no	yes	yes
risk_self			0.30			0.62			0.60			−0.75
			(0.53)			(0.66)			(0.70)			(0.60)
risk_bomb			0.12			0.14			0.16			0.83^{***}
			(0.20)			(0.23)			(0.25)			(0.27)
fair1			2.21^{***}			−1.03			−0.47			0.56
			(0.85)			(1.08)			(1.10)			(0.94)
fair2			−0.61			0.42			0.22			1.89^{*}
			(1.06)			(1.07)			(1.16)			(0.99)
donation			-0.02^{***}			0.01^{*}			0.01			0.01^{*}
			(0.01)			(0.01)			(0.01)			(0.01)
share_self			0.14			-1.16^{**}			0.58			0.91^{*}
			(0.43)			(0.53)			(0.55)			(0.47)
constant	40.8^{***}	52.9^{***}	50.0^{***}	1.08	14.46	13.09	−2.38	−5.06	−12.85	−0.30	7.00	−2.60
	(4.49)	(12.97)	(12.78)	(2.42)	(11.72)	(13.43)	(2.48)	(11.88)	(12.77)	(2.36)	11.25	(10.94)
Adj. R^2	0.82	0.81	0.83	0.76	0.76	0.76	0.71	0.70	0.70	0.82	0.82	0.85
n	104	100	100	104	100	100	112	109	109	112	109	109

[8] $p - value < 0.10 : *; < 0.05 : **; < 0.01 : ***$. And standard errors are in the parentheses.

However, as shown in columns (3), (6), (9), and (12) of Table 5, incorporating social preferences into the model does not meaningfully affect the estimated coefficients on *value* or $value_m$. Furthermore, the signs of the coefficients for social preference variables appear inconsistent across specifications. For example, in "Complete", subjects who report greater willingness to donate after a windfall in the lottery tend to propose lower contributions to the joint project. One possible explanation is that responses to the questionnaire reflect post-hoc rationalization rather than stable preferences, limiting their predictive value for behavior in the bargaining task.

Although some of the behavior observed among Low-type participants in "Dictator"—particularly low proposals and low MACs—is consistent with aversion to unequal outcomes, the regression results do not support a stable or significant role for fairness preferences as measured. This discrepancy likely reflects limitations of post-experimental surveys in capturing context-specific motivations. While some decisions may be driven by inequality aversion, such preferences may not be reliably expressed through general attitudinal questions detached from the strategic environment.

5 Concluding Discussion

This study experimentally investigates whether externally enforced random dictatorship ("Dictator") can mitigate the coordination failures commonly observed in threshold public goods bargaining under incomplete information. Drawing on the theoretical framework of Myerson (1984), the analysis compares three conditions: Complete Information ("Complete"), Incomplete Information ("Incomplete"), and Incomplete Information with Random Dictatorship ("Dictator").

The results confirm that incomplete information significantly reduces the likelihood of agreement, especially when participants differ in their valuations of the joint project. Introducing the Random Dictatorship mechanism improves the alignment between participants' willingness to contribute and their private valuations, suggesting that Random Dictatorship facilitates more truthful preference revelation. However, it does not significantly improve agreement rates relative to “Incomplete”. This limitation appears to stem from a reluctance to contribute when not selected as the proposer, even when doing so would be individually rational.

While the observed behavior in “Dictator” deviates from theoretical predictions, many participants appear to adjust their decisions in ways consistent with “best responses” to their expectations about others’ choices. Notably, high-type subjects tend to offer contributions that would yield positive expected returns, even if lower than the theoretically optimal strategy. For low-type participants, inequality aversion may explain deviations from equilibrium behavior, although this motivation is not strongly reflected in post-experimental measures of social preferences.

These findings highlight both the promise and limits of mechanism-based solutions to coordination under private information. Random dictatorship succeeds in partially addressing incentive alignment and eliciting private values, but fails to overcome residual strategic hesitation in one-shot interactions. The study thereby underscores the need for complementary mechanisms—such as communication, repeated interaction, or stronger institutional enforcement—to support agreement in environments where strategic uncertainty persists.

Future work may explore whether similar results hold in richer environments with more than two players, variable project costs, or endogenous mechanism selection. Moreover, improving the measurement of context-sensitive preferences may help better explain deviations from equilibrium behavior in asymmetric bargaining settings.

References

[1] Lawrence M. Ausubel, Peter Cramton, and Raymond Deneckere. Bargaining with incomplete information. *Handbook of Game Theory with Economic Applications*, 3:1897–1945, 2002.

[2] Matthew Backus, Thomas Blake, Brad Larsen, and Steven Tadelis. Sequential bargaining in the field: Evidence from millions of online bargaining interactions. *The Quarterly Journal of Economics*, 135(3):1319–1361, 2020.

[3] Steven J Brams and Samuel Merrill. Equilibrium strategies for final-offer arbitration: There is no median convergence. *Management Science*, 29(8):927–941, 1983.

[4] Colin F Camerer, Gideon Nave, and Alec Smith. Dynamic unstructured bargaining with private information: theory, experiment, and outcome prediction via machine learning. *Management Science*, 65(4):1867–1890, 2019.

[5] Kalyan Chatterjee and Bhaskar Dutta. Bargaining under incomplete information: the role of time preferences. *Games and Economic Behavior*, 5(1):97–117, 1993.

[6] Kalyan Chatterjee and William Samuelson. Bargaining under incomplete information. *Operations research*, 31(5):835–851, 1983.

[7] In-Koo Cho and David M. Kreps. Signaling games and stable equilibria. *Quarterly Journal of Economics*, 102(2):179–221, 1987.

[8] Cary Deck and Amy Farmer. Bargaining over an uncertain value: Arbitration mechanisms compared. *Journal of Law, Economics, and Organization*, 23(3):547–579, 2007.

[9] David L. Dickinson. Labor negotiations, conflicts, and arbitration. In Klaus F. Zimmermann, editor, *Handbook of Labor, Human Resources and Population Economics*, pages 1–20. Springer Nature, 2020.

[10] Tore Ellingsen, Magnus Johannesson, Jannie Lilja, and Henrik Zetterqvist. Trust and truth. *The Economic Journal*, 119(534):252–276, 2009.

[11] Henry S Farber. An analysis of final-offer arbitration. *Journal of Conflict Resolution*, 24(4):683–705, 1980.

[12] Robert Forsythe, John Kennan, and Barry Sopher. An experimental analysis of strikes in bargaining games with one-sided private information. *The American Economic Review*, pages 253–278, 1991.

[13] Drew Fudenberg and J. Miguel Villas-Boas. Inference of preferences and voluntary disclosure. *Games and Economic Behavior*, 59(1):1–30, 2007.

[14] Maria Jose Herrero and Cesar Martinelli. A model of the strategic foundations of axiomatic bargaining theory. *Games and Economic Behavior*, 24(1-2):286–323, 1998.

[15] Sönke Hoffmann, Benedikt Mihm, and Joachim Weimann. To commit or not to commit? an experimental investigation of pre-commitments in bargaining situations with asymmetric information. *Journal of Public Economics*, 121:95–105, 2015.

[16] Ehud Kalai and Meir Smorodinsky. Other solutions to nash's bargaining problem. *Econometrica*, 43(3):513–518, 1975.

[17] Bradley J Larsen. The efficiency of real-world bargaining: Evidence from wholesale used-auto auctions. *The Review of Economic Studies*, 88(2):851–882, 2021.

[18] Tobias Lundquist, Tore Ellingsen, Erik Gribbe, and Magnus Johannesson. The aversion to lying. *Journal of Economic Behavior & Organization*, 70(1-2):81–92, 2009.

[19] Michael Mitzkewitz and Rosemarie Nagel. Experimental results on ultimatum games with incomplete information. *International Journal of Game Theory*, 22(2):171–198, 1993.

[20] Roger B. Myerson. Two-person bargaining problems with incomplete information. *Econometrica*, 52(2):461–487, 1984.

[21] John F. Nash. The bargaining problem. *Econometrica*, 18(2):155–162, 1950.

[22] Roy Radner and Andrew Schotter. The sealed-bid mechanism: An experimental study. *Journal of Economic Theory*, 48(1):179–220, 1989.

[23] Amnon Rapoport, Terry E Daniel, and Darryl A Seale. Reinforcement-based adaptive learning in asymmetric two-person bargaining with incomplete information. *Experimental Economics*, 1(3):221–253, 1998.

[24] Kathleen Valley, Leigh Thompson, Robert Gibbons, and Max H Bazerman. How communication improves efficiency in bargaining games. *Games and Economic Behavior*, 38(1):127–155, 2002.

Appendix

Appendix A

Appendix A: Instructions

Instructions for Random Dictatorship treatment

Welcome to this study on economic decision-making.

The study consists of two phases:

- in **Phase 1** you will be matched with another participant and make some decisions.
- in **Phase 2** you will have to fill out a questionnaire.

At the end of the study, you will receive **an amount of money** which depends on your decisions and on the decisions of the other participant, **plus** a participation fee of **1 euro**.

-------Page Break-------

PLEASE read the following instructions carefully.

During instructions, you will have to answer several control questions, and **only after you answer them all correctly**, you can **move on** to the next part.

-------Page Break--------

Instructions for Phase 1

In this phase, you will be **randomly paired** with another participant, whom we will refer to as "**your match**".

What should you do?

You and your match can invest in a **joint project**.

The joint project **has a total cost of 100 tokens**, which you and your match must cover with a joint investment.

The **value** of the project for **you** is **X** *(X= 30 or 90)* tokens, while the **value** of the project for **your match** could be **30 or 90 tokens with equal probability**.

If X=30, we display:

- If the value of the project for **your match is 30 tokens**, then the project is **not worth** the investment since it costs 100 tokens and generates only 60 tokens (30 for you and 30 for your match).

- If instead the value of the project for **your match is 90 tokens**, then the project is **worth** the investment since it costs 100 tokens and generates a total value of 120 tokens (30 for you and 90 for your match).

If X=90, we display:

Regardless of the value of this project for your match, the project is **worth** the investment since it costs 100 tokens and generates a total value of either 120 tokens (90 for you and 30 for your match) or 180 tokens (90 for you and 90 for your match).

However, you and your match must decide how to split the cost, without knowing each other's value. Your match knows their own value, but does not know yours, and is only informed that your value could be either 30 or 90 with equal probability.

[*No previous button*]

--- Page Break ----

Control questions:

1. What is **the value of the project for you**?
    a. 30 tokens
    b. 90 tokens
    c. 100 tokens
    d. 30 tokens or 90 tokens, each with a 50% probability

[*correct answer: a if X=30; b if X=90*]

2. What is **the value of the project for your match**?
    a. 30 tokens
    b. 90 tokens
    c. 100 tokens
    d. 30 tokens or 90 tokens, each with a 50% probability

[correct answer: d]

3. What is the **total cost of the project**?
    a. 30 tokens
    b. 90 tokens
    c. 100 tokens
    d. 30 tokens or 90 tokens, each with a 50% probability

[*correct answer: c*]

[*They can continue only after they answer all the questions correctly. If they make mistakes, they will be sent back to the instruction and answer the questions again.]*

----Page Break----

Fail_cq1

You have made mistake(s) in answering the control questions.

You'll be sent back to the instructions after clicking the button.

Please read the instructions carefully and answer these questions again.

[*No previous button*]

----Page Break----

Succeed_cq1

You have answered all these control questions correctly.

Please click the button to go to the next part.

[*No previous button*]

----Page Break----

How do you and your match decide how to split the cost of the project?

The computer will **randomly select** either you or your match to make a proposal.

<u>If you are selected</u>, you will have to propose **how much you contribute** to cover the cost of the common project. The rest of the cost should be covered by your match.

You can propose **any number between 0 and your value X**.

Your match will then decide whether to **accept or reject** your proposal.

1. If your match **accepts** it, the project is realized, and **you earn X minus your contribution**. Your match earns their value for the project minus their contribution.
2. If instead your match **rejects** your proposal, the project is not realized, and **you and your match both earn 0**.

Remember that both you and your match will make this decision **without knowing the value of this project to each other**.

-----Page Break----

Control questions:

1. If you are selected, you propose to contribute X/3 tokens, how much will you earn?
    a. X/3 tokens if your match accepts your proposal; 0 if your match rejects your proposal
    b. X/3 tokens if your match rejects your proposal
    c. 0
    d. 2X/3 tokens if your match accepts your proposal

[*correct answer: d*]

2. If you are selected, you propose to contribute X tokens, how much will you earn?
    a. X/3 tokens if your match accepts your proposal; 0 if your match rejects your proposal
    b. X/3 tokens if your match rejects your proposal
    c. 0
    d. 2X/3 tokens if your match accepts your proposal

[*correct answer: c*]

----Page Break----

What if the computer selects your match to make a proposal?

<u>If you are NOT selected</u>, your match will make a proposal and **you will have to either accept or reject** it.

3. If you **accept** it, the project is realized, and **you earn X minus your contribution**. Your match earns their value for the project minus their contribution.
4. If instead you **reject** your match's proposal, the project is not realized, and **you and your match both earn 0**.

----Page Break----

Control questions:

1. Suppose you are not selected by the computer, and your match proposes to contribute 100-X/2 tokens, how much will you earn?
    a. X/2 tokens if you accept your match's proposal; 0 if you reject your match's proposal
    b. X/2+10 tokens if you accept your match's proposal
    c. X tokens
    d. 0

[*correct answer: a*]

2. Suppose you are not selected by the computer, and your match proposes to contribute 100-X tokens, how much will you earn?

a. 10 if you accept your match's proposal; 0 if you reject your match's proposal
b. X tokens if you accept your match's proposal
c. 2X/3 tokens if you accept your match's proposal
d. 0

[*correct answer: d*]

-----Page Break-----

We will ask you to **make your choices before knowing** whether you or your match will be **selected** by the computer to make the proposal. Hence, you will have to make **2 decisions**.

First, you will **input your proposed contribution** in case you are selected. Remember that you have to make a proposal without knowing whether the value of the project for your match is 30 or 90 tokens.

Second, you will have to **decide whether you would accept or reject your match's proposal** in case you are not selected by the computer.

-----Page Break-----

If you are not selected, you must decide whether to accept or reject a proposal without knowing the exact amount your match proposed. We will ask you **a sequence of questions to determine the minimum proposal you are willing to accept**. For example

1. Would you accept a proposal of 10 tokens, which implies that you pay 90 tokens for the joint project?
2. Would you accept a proposal of 90 tokens, which implies that you pay 10 tokens for the joint project?

Then if you are not selected, we will match your answers with your match's proposal to determine whether you accept or reject it.

----Page Break----

Earnings

At the end of this study, your final earnings in this phase will be determined, according to the following exchange rate, **5 tokens = 1 euro.**

----Page Break----

You have read all the instructions for phase 1. Here you can have a review over the complete instruction file by clicking the link below.

Link to the instruction file

[*check box:*

You have read all the instructions for phase 1. Do you confirm that you fully understand what you should do and how your payoff will be determined by your decision making?

- *I confirm* - *Back to the Instructions*

]

----Page Break----

Proposal

The joint project **requires** a total investment of **100 tokens**.

The **value** of the project for **you** is **X tokens**.

The **value** of the project for **your match** is **30 or 90 tokens, each with a 50% probability**.

Please choose **your proposed contribution** (integers only):

[slider from 0 to X]

[button with check box]: If you are selected and your proposal is accepted, you earn X-your contribution tokens *[in value]*. If your proposal is rejected, you earn 0. Do you confirm your choice?

-----Page Break-----

Second player

Your proposal has been recorded.

Now consider the case in which you are not selected by the computer.

We will ask you **a sequence of questions** (about 4 to 7 questions) to determine the minimum proposal you are willing to accept.

------Page Break------

Questions *[Y changes in different questions]*

The joint project **requires** a total investment of **100 tokens.**

The **value** of the project for **you** is **X** tokens.

The **value** of the project for **your match** is **30 or 90 tokens, each with a 50% probability**.

If you are **not selected** by the computer, and **your match proposes to contribute Y tokens**, that is, **you** will have to **contribute 100-Y tokens** to this joint project, would you accept or reject your match's proposal?

- Accept
- Reject

[Multiple choice]

--------Page Break------

[*Questionnaire*]

-----Page Break-------

Thanks for your time spent taking this survey.

Your response has been recorded.

We'll inform you of the bargaining outcome later through a Prolific message.

After clicking the "Complete" button, you'll be redirected to Prolific's "completed" page.

Instructions for Incomplete Information treatment

Welcome to this study on economic decision-making.

The study consists of two phases:

- in **Phase 1** you will be matched with another participant and make some decisions.
- in **Phase 2** you will have to fill out a questionnaire.

At the end of the study, you will receive **an amount of money** which depends on your decisions and on the decisions of the other participant, **plus** a participation fee of **1 euro**.

-------Page Break-------

PLEASE read the following instructions carefully.

During instructions, you will have to answer several control questions, and **only after you answer them all correctly**, you can **move on** to the next part.

-------Page Break--------

Phase 1

In this phase, you will be **randomly paired** with another participant, whom we will refer to as "**your match**".

What should you do?

You and your match can invest in a **joint project**.

The joint project **has a total cost of 100 tokens**, which you and your match must cover with a joint investment.

The **value** of the project for **you** is **X** *(X= 30 or 90)* tokens, while the **value** of the project for **your match** could be **30 or 90 tokens with equal probability**.

If X=30, we display:

- If the value of the project for **your match is 30 tokens**, then the project is **not worth the investment** since it costs 100 tokens and generates only 60 tokens (30 for you and 30 for your match).
- If instead the value of the project for **your match is 90 tokens**, then the project is **worth** the investment since it costs 100 tokens and generates a total value of 120 tokens (30 for you and 90 for your match).

If X=90, we display:

Regardless of the value of this project for your match, the project is **worth** the investment since it costs 100 tokens and generates a total value of either 120 tokens (90 for you and 30 for your match) or 180 tokens (90 for you and 90 for your match).

However, you and your match must decide how to split the cost, without knowing each other's value. Your match knows their own value, but does not know yours, and is only informed that your value could be either 30 or 90 with equal probability.

--- Page Break ----

Control questions:

1. What is **the value of the project for you**?
   a. 30 tokens
   b. 90 tokens
   c. 100 tokens
   d. 30 tokens or 90 tokens, each with a 50% probability

[*correct answer: a if X=30; b if X=90*]

2. What is **the value of the project for your match**?
   a. 30 tokens
   b. 90 tokens
   c. 100 tokens
   d. 30 tokens or 90 tokens, each with a 50% probability

[correct answer: d]

3. What is the **total cost of the project**?
   a. 30 tokens
   b. 90 tokens
   c. 100 tokens
   d. 30 tokens or 90 tokens, each with a 50% probability

[*correct answer: c*]

[*They can continue only after they answer all the questions correctly. If they make mistakes, they will be sent back to the instruction and answer the questions again.]*

----Page Break----

How do you and your match decide how to split the cost of the project?

You will have to propose **how much you contribute** to cover the cost of the common project. The rest of the cost should be covered by your match.

You can propose **any number between 0 and your value X**.

Your match will have to **do the same**.

Remember that both you and your match will make the proposal **without knowing the value of this project to each other**.

The outcomes depend on **<u>the sum of your and your match's proposed contributions</u>**:

- If **this sum is less than 100 tokens**, the project is **not realized**, and **both** you and your match will **earn 0**
- If **this sum is larger than or equal to 100 tokens**, the project **is realized**, and the computer will **randomly implement** one of the two proposals
  - If **your proposal is implemented**, **you'll contribute as you proposed**, and your match will cover the rest of the investment. Thus, **you will earn X minus your proposed contribution**, and your match will earn their value minus the rest of the investment needed.
  - If **your match's proposal is implemented**, your match will contribute as they proposed, and **you will cover the rest of the investment**. Thus, **you will earn X minus your contribution**, and your match will earn their value minus their proposed contribution.

-----Page Break----

Control questions:

1. Suppose you propose to contribute 2X/3 tokens, and your match proposes to contribute 80-2X/3 tokens, how much will you earn?
   a. X/3 tokens
   b. X/3-20 tokens
   c. 0
   d. X/3 tokens or X/3-20 tokens, each with a 50% probability

[correct answer: c]

2. Suppose you propose to contribute 2X/3 tokens, and your match proposes to contribute 110-2X/3 tokens, how much will you earn?
   a. X/3 tokens if your proposal is implemented
   b. X/3+20 tokens or X/3+10 tokens, each with a 50% probability
   c. 0
   d. X/3 if your match's proposal is implemented

[*correct answer: a*]

3. Suppose you propose to contribute 2X/3 tokens, and your match proposes to contribute 100-2X/3 tokens, how much will you earn?
    a. 0
    b. 2X/3 tokens if your proposal is implemented
    c. X/3 tokens
    d. 2X/3 tokens if your match's proposal is implemented

[*correct answer: c*]

----Page Break----

Earnings

At the end of this study, your final earnings in this phase will be determined, according to the following exchange rate, **5 tokens = 1 euro.**

----Page Break----

Proposal

The joint project **requires** a total investment of **100 tokens**.

The **value** of the project for **you** is **X tokens**.

The **value** of the project for **your match** is **30 or 90 tokens, each with a 50% probability**.

Please choose **your proposed contribution** (integers only):

[slider from 0 to X]

[check box:

- *If the sum of your and your match's proposed amount is no less than 100, and if your proposal is implemented, you earn <number> tokens.*

- *If the sum of your and your match's proposed amount is less than 100, you earn 0.*

Do you confirm your choice?]

-----Page Break-----

[*Questionnaire*]

-----Page Break-------

Thanks for your time spent taking this survey.

Your response has been recorded.

We'll inform you of the bargaining outcome later through a Prolific message.

After clicking the "Complete" button, you'll be redirected to Prolific's "completed" page.

Instructions for Complete Information treatment

Welcome to this study on economic decision-making.

The study consists of two phases:

- in **Phase 1** you will be matched with another participant and make some decisions.
- in **Phase 2** you will have to fill out a questionnaire.

At the end of the study, you will receive **an amount of money** which depends on your decisions and on the decisions of the other participant, **plus** a participation fee of **1 euro**.

-------Page Break-------

PLEASE read the following instructions carefully.

During instructions, you will have to answer several control questions, and **only after you answer them all correctly**, you can **move on** to the next part.

-------Page Break--------

Phase 1

In this phase, you will be **randomly paired** with another participant, whom we will refer to as "**your match**".

What should you do?

You and your match can invest in a **joint project**.

The joint project **has a total cost of 100 tokens**, which you and your match must cover with a joint investment.

The **value** of the project for **you** is **X** *(X= 30 or 90)* tokens, while the **value** of the project for **your match** is **Y** *(Y= 30 or 90)* tokens. The project is **worth** the investment since it costs 100 tokens and generates a total value of X+Y tokens (X for you and Y for your match).

[The case when X=Y=30 will be eliminated]

However, you and your match must decide how to split the cost, knowing the value of the project to each other.

--- Page Break ----

Control questions:

1. What is **the value of the project for you**?

a. 30 tokens
b. 90 tokens
c. 100 tokens
d. 30 tokens or 90 tokens, each with a 50% probability

[*correct answer: a if X=30; b if X=90*]

2. What is **the value of the project for your match**?
    a. 30 tokens
    b. 90 tokens
    c. 100 tokens
    d. 30 tokens or 90 tokens, each with a 50% probability

[correct answer: a if Y=30; b if Y=90]

3. What is the **total cost of the project**?
    a. 30 tokens
    b. 90 tokens
    c. 100 tokens
    d. 30 tokens or 90 tokens, each with a 50% probability

[*correct answer: c*]

[*They can continue only after they answer all the questions correctly. If they make mistakes, they will be sent back to the instruction and answer the questions again.]*

----Page Break----

How do you and your match decide how to split the cost of the project?

You will have to propose **how much you contribute** to cover the cost of the common project. The rest of the cost should be covered by your match.

You can propose **any number between 0 and your value X**.

Your match will have to **do the same**.

The outcomes depend on **the sum of your and your match's proposed contribution**:

- If **this sum is less than 100 tokens**, the project is **not realized**, and **both** you and your match will **earn 0**
- If **this sum is larger than or equal to 100 tokens**, the project **is realized**, and the computer will **randomly implement** one of the two proposals
    - If **your proposal is implemented**, **you'll contribute as you proposed**, and your match will cover the rest of investment. Thus, **you will earn X minus your proposed contribution**, and your match will earn Y minus their contribution.

- If **your match's proposal is implemented**, your match will contribute as they proposed, and **you will cover the rest of investment**. Thus, **you will earn X minus your contribution**, and your match will earn Y minus their proposed contribution.

-----Page Break----

Control questions:

1. Suppose you propose to contribute (X-Y+80)/2 tokens, and your match proposes to contribute (Y-X+80)/2 tokens, how much will you earn?
   a. (X+Y-80)/2 tokens
   b. (X+Y-80)/2-10 tokens
   c. 0
   d. (X+Y-80)/2 tokens or (X+Y-80)/2-10 tokens, each with a 50% probability

[correct answer: c]

2. Suppose you propose to contribute (X-Y+110)/2 tokens, and your match proposes to contribute (Y-X+110)/2 tokens, how much will you earn?
   a. (X+Y-110)/2 tokens if your proposal is implemented
   b. (X+Y-100)/2 tokens or (X+Y-90)/2 tokens, each with a 50% probability
   c. 0
   d. (X+Y-110)/2 if your match's proposal is implemented

[*correct answer: a*]

3. Suppose you propose to contribute (X-Y+100)/2 tokens, and your match proposes to contribute (Y-X+100)/2 tokens, how much will you earn?
   a. 0
   b. (X+Y-80)/2 tokens if your proposal is implemented
   c. (X+Y-100)/2 tokens
   d. (X+Y-80)/2 tokens if your match's proposal is implemented

[*correct answer: c*]

----Page Break----

Earnings

At the end of this study, your final earnings in this phase will be determined, according to the following exchange rate, **5 tokens = 1 euro.**

----Page Break----

Proposal

The joint project **requires** a total investment of **100 tokens**.

The **value** of the project for **you** is **X tokens**.

The **value** of the project for **your match** is **Y tokens.**

Please choose **your proposed contribution** (integers only):

[slider from 0 to X]

-----Page Break-----

[*Questionnaire*]

-----Page Break-------

Thanks for your time spent taking this survey.

Your response has been recorded.

We'll inform you of the bargaining outcome later through a Prolific message.

After clicking the "Complete" button, you'll be redirected to Prolific's "completed" page.

Appendix B

Appendix B: Questionnaire

Questionnaire

1. What is your age?

[*slider: value between 0 and 99*]

2. What is your gender?

[*multiple choice: gender*]

Female / Male / Non-Binary / Prefer not to say

3. What is your Major?

1) Please indicate the macro thematic area:

[*multiple choice: macro area]*

Math, Physics | Econ, Stats | Business & Management | Biology, Chemistry & other Natural sciences | Engineering & Architecture | Languages | Classical Studies | Arts | Psychology, Sociology & other social sciences | Law & Political Sciences | Medicine & related disciplines | Other

2) Please specify what's your specific field:

[*text entry*]

4. What is your nationality?

Please, first select the continent.

[*multiple choice: nationality*]

Africa / Antarctica / Asia / Central and South America / Europe / North America / Oceania

Please specify your nationality by filling in the blank below.

[*text entry*]

5. Please tell me, in general, how willing or unwilling you are to take risks. Please use a scale from 0 to 10, where 0 means "completely unwilling to take risks" and a 10 means you are "very willing to take risks". You can also use any numbers between 0 and 10 to indicate where you fall on the scale, like 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10

[*scale from 0 to 10*]

6. Below you can see a field composed of 5*5=25 boxes. Behind one of these boxes a bomb is hidden; the remaining 24 boxes are empty. You do not know where the bomb is. You only know that it can be in any place with equal probability. Your task is to choose how many boxes to collect. Boxes will be collected in numerical order, so you will be asked to choose a number between 1 and 25.

If you happen to have collected the box in which the time bomb is located – i.e., if your chosen number is greater than, or equal to, the number of the box with bomb – you will earn zero. If the time bomb is located in a box that you did not collect – i.e., if your chosen number is smaller than the number of the box with bomb – you will earn an amount in euro equivalent to the number you have chosen divided by ten.

[*numeric entry*]

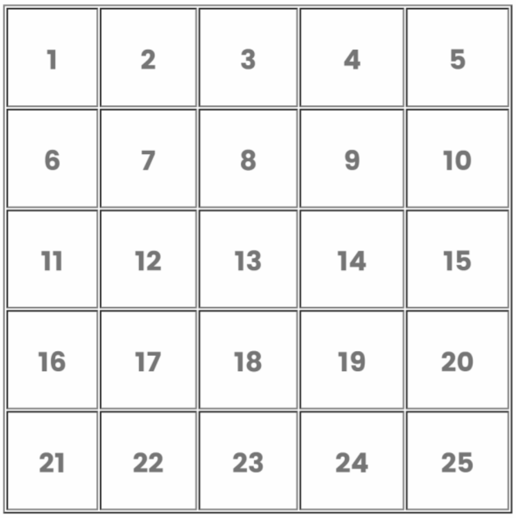


Figure 3 25 Boxes

7. To what extent do you agree with the following statement: “It is unfair if luck determines people’s economic situation.”

[multiple choice: fairness_preference_1]

Scale of Strongly disagree/ Somewhat disagree/ Neither agree nor disagree/ Somewhat agree/ Strongly agree

8. To what extent do you agree with the following statement: “Luck is an important determinant of people’s economic situation.”

[multiple choice: fairness_preference_2]

Scale of Strongly disagree/ Somewhat disagree/ Neither agree nor disagree/ Somewhat agree/ Strongly agree.

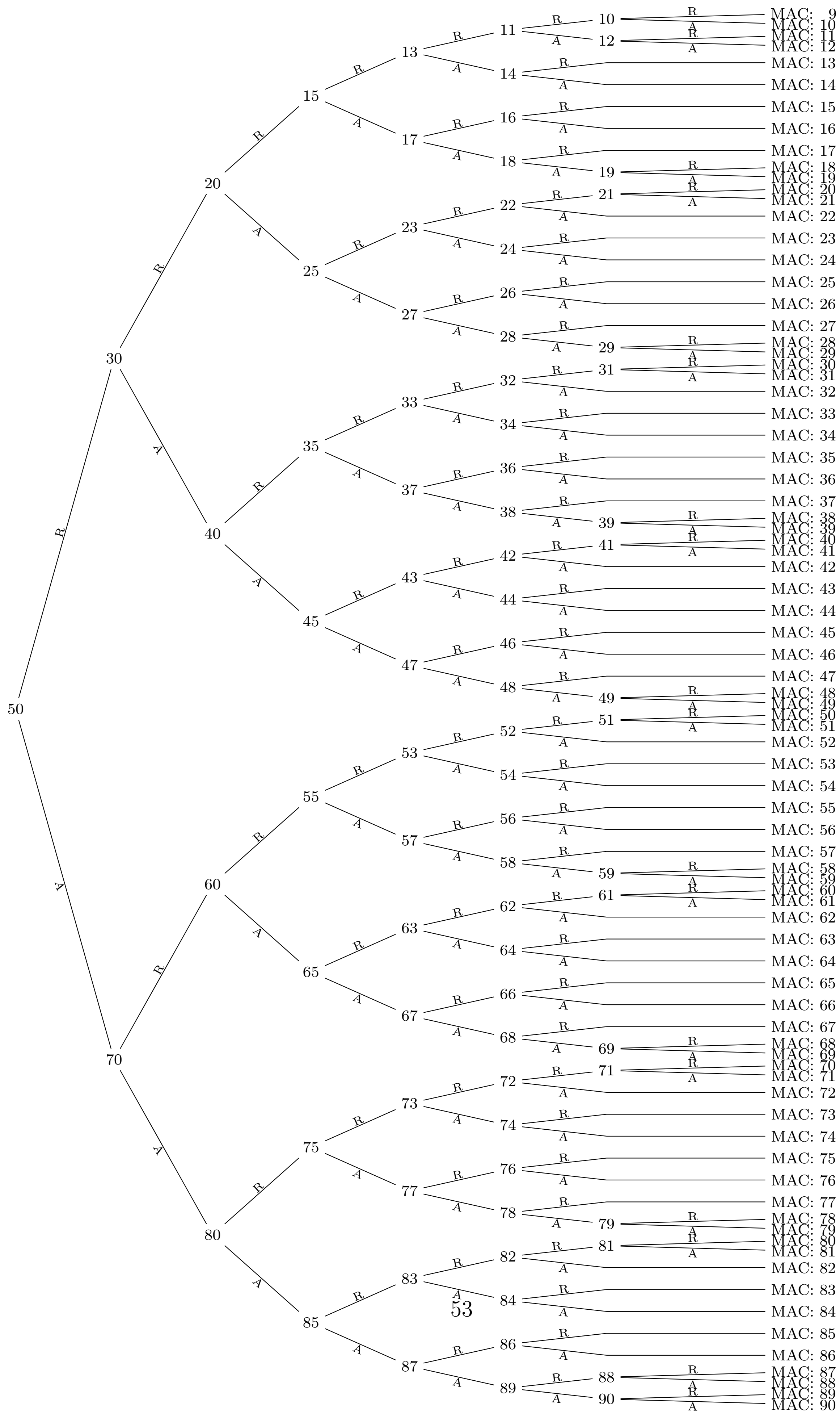
9. Imagine the following situation: you won 1,000 Euro in a lottery. Considering your current situation, how much would you donate to charity?

[*slider: value between 0 to 1000*]

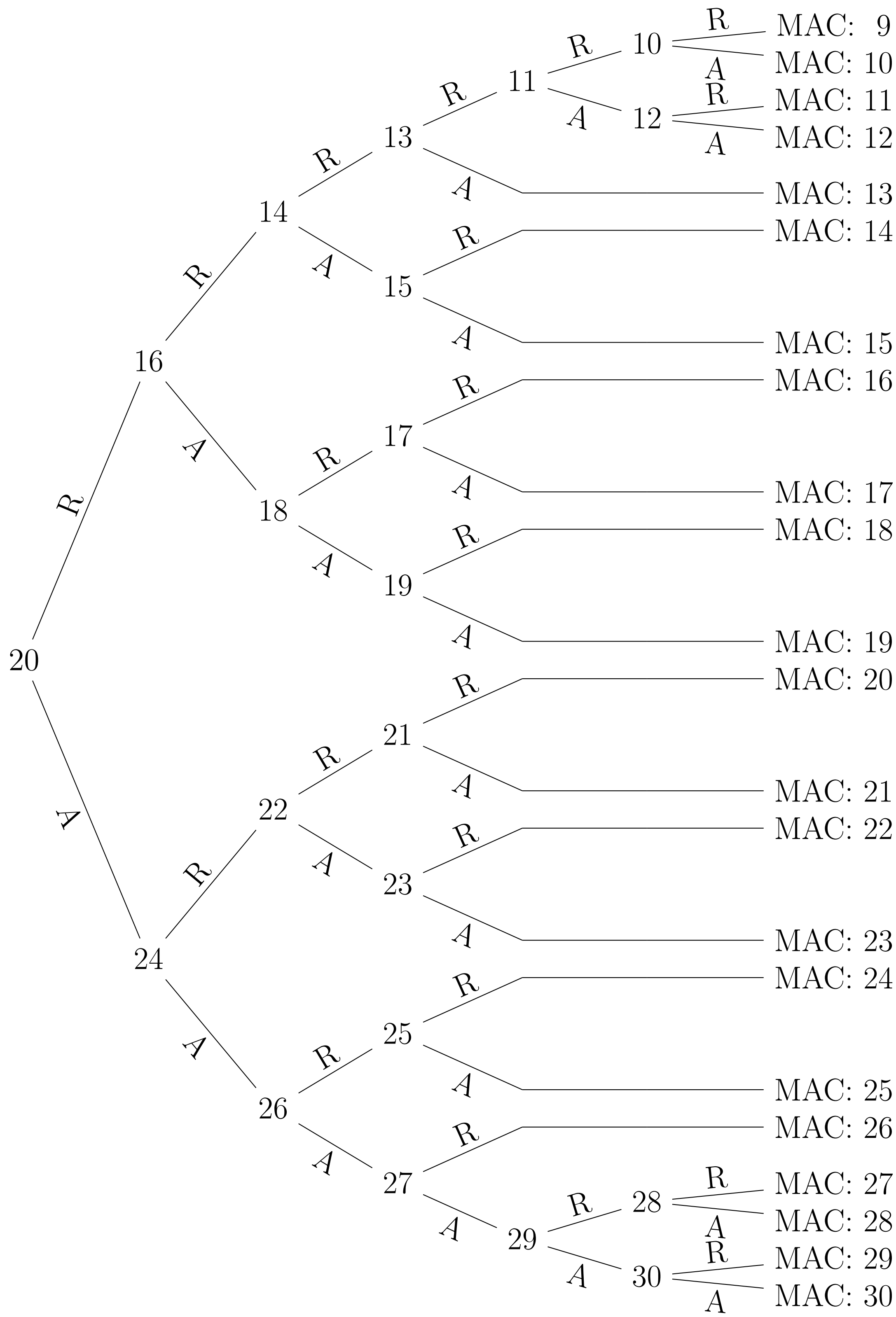
10. How do you assess your willingness to share with others without expecting anything in return when it comes to charity? Please use a scale from 0 to 10, where 0 means you are “completely unwilling to share" and a 10 means you are “very willing to share". You can also use the values in-between to indicate where you fall on the scale.

[*slider: value between 0 to 10*]

Appendix C.1 Staircase Procedures for High-type subjects



Appendix C.2 Staircase Procedures for Low-type subjects



Appendix D Homogeneity of Subjects Samples across Treatments

Table 6: Summary Statistics of Demographics and Social Preferences across Treatments

	Complete			Incomplete			Dictator			$p-value$[9]		
	n	mean	sd	n	mean	sd	n	mean	sd	C[10] vs. I	C vs. D	I vs. D
age	104	23.54	1.82	104	23.22	2.39	112	23.22	2.28		> 0.2	
gender[11]	100	0.65	0.48	100	0.60	0.49	109	0.63	0.48		> 0.2	
risk_self[12]	104	5.22	1.80	104	5.52	1.78	112	5.45	1.88		> 0.2	
risk_bomb[13]	104	12.90	4.57	104	11.93	5.21	112	12.25	4.78		> 0.15	
fair1[14]	104	0.83	1.09	104	0.93	1.09	112	0.52	1.09	0.42	0.02	0.00
fair2[15]	104	0.78	0.87	104	0.57	1.07	112	0.60	1.04		> 0.15	
donation [16]	104	164.82	171.84	104	174.99	157.62	112	141.04	154.45	0.40	0.53	0.07
WTS_self [17]	104	6.17	2.08	104	6.61	2.18	112	6.03	2.31	0.10	0.63	0.04

As is shown above, the subject samples are homogeneous across treatments except: 1. subjects in "Dictator" significantly less agree with "It is unfair if luck determines people's economic situation."; 2. subjects in "Dictator" are willing to donate significantly less money to a good cause if they win 1000 euros in the lottery; 3. subjects in "Complete" and "Dictator" self-evaluated themselves as significantly less willing to share with others. I will try to figure out the reason for such a discrepancy and do a robustness check in the future.

[9] All the $p-value$s are calculated by Mann-Whitney U test.

[10] "C" represents "Complete", "I" represents "Incomplete", and "D" represents "Dictator".

[11] Male= 0, Female=1, so the mean value means the share of females in the sample.

[12] This represents the answers to Question 5 collecting the self-evaluated risk-loving level in the Questionnaire (Appendix B). The higher the value is, the more risk-loving the subject is.

[13] This represents subjects' choice in the bomb revelation task for risk elicitation (Question 6). The higher the number is, the more risk-loving the subject is.

[14] This represents the subjects' attitude towards "It is unfair if luck determines people's economic situation." (Question 7). The higher the value is, the more the subjects agree with this statement.

[15] This represents the subjects' attitude towards "Luck is an important determinant of people's economic situation." (Question 8). The higher the value is, the more the subjects agree with this statement.

[16] This reveals subjects' self-evaluated donation amount when they win 1000 euros in the lottery (Question 9). The higher the value is, the more altruistic the subject is.

[17] This is subjects' self-evaluated willingness to share (Question 10). The higher the value is, the more altruistic the subject is.

Maria Bigoni[1] – Stefania Bortolotti[2] – Francesco Parisi[3] – Xin Zhang[4]

Do Contract Remedies Affect Efficient Renegotiation? An Experiment

ABSTRACT: Rational parties enter into a contract if the agreement is mutually beneficial. However, after the contract is formed, changes to the costs and/or benefits of performance may render the original contract undesirable. In this paper, we carry out an incentivized experiment to study the effect of alternative remedies on the parties' ability to renegotiate their contractual obligations. After entering into a contract, experimental subjects observe symmetrical changes to the original costs and/or benefits, which create a misalignment of their performance vs. breach incentives. Our experimental design compares the effects of damage and specific performance remedies on the parties' ability to renegotiate. In the overwhelming majority of cases in which a shock occurs, participants enter into renegotiation and successfully conclude it. We find no treatment differences in terms of renegotiation entry/success or surplus division; these results are consistent with Coase's (1960) irrelevance of remedies proposition in this streamlined setting.

Keywords: Efficient breach, contract remedies, specific performance, damages

JEL Codes: K12, K41, C90, D86, D91

1. Introduction

Whether one legal remedy is more efficient than another in fostering an efficient contract breach is ultimately an empirical issue. This paper investigates the micro-determinants of renegotiation under alternative legal remedies for breaches of contracts. Specifically, we aim to determine whether the choice of legal remedy—specific performance or damages—affects contracting parties' willingness and ability to renegotiate in the wake of an exogenous market shock to prevent an inefficient breach or an inefficient performance of a contract.

We consider two symmetrical contractual scenarios involving a sale contract. In the first scenario, the contract is subject to a damages remedy. The promisor (seller) has a credible threat to breach, and the promisee (buyer) would need to negotiate to prevent the breach and receive the good or service that was promised. In the second scenario, the contract is subject to

[1] Professor of Economics at the University of Bologna, Department of Economics (Corresponding author. Email: maria.bigoni@unibo.it).
[2] Associate Professor of Economics at the University of Bologna, Department of Economics (Email: stefania.bortolotti@unibo.it).
[3] Oppenheimer Wolff and Donnelly Professor of Law at the University of Minnesota, and Professor of Economics at the University of Bologna, Department of Sociology and Business Law. Visiting Professor of Law, University of Miami (Email: parisi@umn.edu).
[4] Ph.D. (Economics) Candidate, University of Bologna, Department of Economics (Email: xin.zhang4@unibo.it).

specific performance. The promisee (buyer) has a right to enforce performance and the promisor (seller) would need to negotiate to avoid performance and permissively breach. In each scenario, the contracting parties face financially symmetrical changes to the costs and/or benefits of the original contract after that contract has been formed. This creates a misalignment of their incentives for performance vs. breach. Under each scenario, parties could benefit from the renegotiation of their agreement. With these scenarios in mind, our incentivized experiment assesses whether legal remedies affect contracting parties' willingness and ability to renegotiate to avoid an inefficient breach or an inefficient performance in the face of an exogenous market shock.

Our pre-registered experiment aims to identify the determinants of contract renegotiation in the face of breach opportunities against the theoretical backdrop of Coase's (1960) irrelevance of remedies proposition under idealized, stylized conditions where rationality is most likely to prevail. In a lab experiment with a student sample, we utilize a modified "Contract-Breach Game" à la Bigoni et al. (2017) to elicit micro-level behaviour and to enable causal identification of the effects of contract remedies—these effects would otherwise be difficult to observe or estimate with survey/observational data at the market level. Buyers and sellers must reach an agreement to execute a contract at a predetermined price. Once the contract is signed, an external shock may occur, altering the initial conditions. In the presence of such shocks, enforcing specific performance (under the Specific Performance treatment) or permitting contract breach (under the Damages treatment) leads to aggregate inefficiency, making renegotiation desirable. Should parties opt to renegotiate, they are given 60 seconds to reach a new agreement—either on the compensation owed to the buyer (in the Specific Performance treatment) or on the revised price (in the Damages treatment).

We carefully designed the experiment to minimize framing and demand effects that might otherwise artificially inflate treatment differences, thereby ensuring that any observed behavioural change can be attributed to variation in the underlying economic conditions. In this sense, our setting provides a benchmark test of the irrelevance proposition, conducted under stylized but ideal conditions where rational renegotiation is most likely to occur. Follow-up research can build on this foundation to examine whether the equivalence of remedies breaks down in more complex environments—e.g., when contracts involve reliance investments, legal ambiguity, or emotionally charged entitlements.

Within our set-up, we test how different breach remedies affect parties' (1) willingness to renegotiate a contract, (2) ability to bring renegotiation to a successful completion, and (3) approach to splitting the renegotiation surplus. Overall, 94% of contracting parties proceeded to the renegotiation stage following a shock to the original conditions. Of those, 85% successfully reached an agreement. For both outcomes, we fail to find statistically significant differences between the two treatments. As a result of renegotiation, sellers captured only a small portion of the available surplus—15% in the Damages treatment and 18% in the Specific Performance treatment—again showing no statistically significant effect of remedies on surplus division. More broadly, we observe a strong tendency toward convergence with the predicted equilibria.

As outlined in our pre-registration, we conduct a heterogeneity analysis to assess the role of other-regarding preferences in breach renegotiations. As anticipated, inequality aversion has little influence on the decision to enter renegotiation. However, consistent with prior findings on the impact of individual preferences in bargaining, other-regarding preferences affect the surplus division. In groups where inequality-averse buyers are the majority, they tend to allocate a larger share of the surplus to sellers. We find some—albeit preliminary—evidence that individual predispositions are more likely to influence surplus allocation under the Damages treatment than under Specific Performance.

Our results are consistent with Coase's (1960) irrelevance of remedies proposition: within the limits of our experimental design, we do not find evidence that the choice of remedy affects how contracting parties react to market shocks and their willingness and ability to carry out optimal renegotiation. However, our findings only partially align with the broader predictions of the Coase theorem. While the theorem anticipates universal participation and success in renegotiation, our experiment shows that only 94% of parties initiate renegotiation, and just 85% successfully reach a new agreement. Thus, despite the irrelevance of remedies, only about 80% of subjects fully conform to the rational bargaining expectations set forth by the Coase theorem. Of course, our evidence comes from an intentionally highly controlled and stylized environment, where parties are able to enforce remedies without transaction costs, buyers have not made reliance investments, and the parties cannot hedge against the risk of a shock. While future research should further examine the role of these and other factors in optimal renegotiation under varying remedies, establishing an empirical benchmark and identifying the conditions under which theoretical predictions hold remains essential.

The remainder of the paper is organized as follows. In Section 1.1, we introduce our research question, providing a brief survey of the legal instruments available to enforce contractual promises. In Section 1.2, we identify how our contribution relates to the existing literature. In Section 2, we present our experiment and in Section 3, we discuss our results. In Section 4, we examine the implications of our results for the design of legal frameworks that support optimal contract execution and renegotiation. In Section 5, we discuss the limitations of our experimental design and outline directions for future research. In Section 6, we conclude by summarizing our main findings.

1.1 Research Question

When parties enter into a contract, they expect their counterparts to do as they promised. Yet, contractual promises are not always fulfilled. What should be the legal remedy for a breach of the agreement? One point is clear in all contemporary legal systems: the disappointed promisee should not take the law into her own hands and force the breaching promisor to fulfill his promise. The promisee must instead seek relief by bringing her claims in court. Apart from this common foundation, modern legal systems vary in the types of solutions that they offer when a contractual promise is breached.

Broadly speaking, the promisee's rights can be protected by either a "strong" remedy (where the promisee can force the breaching promisor to fulfill the contractual obligation) or by a "weak" remedy (where the promisor can unilaterally avoid performance by paying damages to his promisee). In the legal terminology, strong remedies are called "specific performance" remedies, whereas weak remedies fall under the general category of "damages." Law and economics scholars (Posner, 1977; Cooter & Ulen, 1997; Edlin, 1998) describe the main characteristics of these two categories of remedies as follows:

1. Specific Performance: In the event of a breach, the promisee can force the promisor to fulfill the contractual obligation and carry out the performance as stated in the contract, *unless* both parties reach an agreement to resolve the contract with the payment of a mutually agreed-upon sum. Under Specific Performance, the promisor can avoid the performance of the original contract *only with* the consent of the promisee and renegotiation of the original contract. By using Calabresi and Melamed's (1972) terminology, this form of relief corresponds to a "property rule."
2. Damages: A promisor is entitled to breach the contract without the consent of his promisee. In the event of a breach, the promisor must pay an amount of damages to his promisee. Damages may be agreed upon by the parties at the time of the contract (liquidated damages) or liquidated by the court to compensate the promisee for the foregone performance benefit (compensatory damages). Under a damages remedy, a promisee who wants to obtain the promised contractual performance, rather than

damages, would have to renegotiate the contract to reverse the promisor's decision to breach. Using Calabresi and Melamed's (1972) terminology, this "weak" remedy corresponds to a "liability rule."

Contemporary legal systems diverge greatly in their use of contract remedies. In the interest of simplicity, we will provide an admittedly stylized mapping of the three main legal approaches, classifying them as (i) strong-remedy systems; (ii) mixed-remedy systems; and (iii) weak-remedy systems.

(i) *Strong-Remedy Systems.* At one end of the spectrum, German law (and other German-based legal systems) adopts strong remedies across nearly the entire range of contractual obligations. Under German law, the legal dogma asserts that a contract grants the promisee a right to obtain performance, not a right to obtain damages in lieu of performance. Section 241 of the German Civil Code (BGB) makes this point clear by stating that a contract gives the creditor a "right to demand performance from the debtor." With a few exceptions, German courts apply this provision, granting specific performance relief for the enforcement of contractual obligations (Treitel, 1988; Zweigert & Kotz, 1998, p. 472-3).

(ii) *Mixed-Remedy Systems.* French law (and other French-based legal systems) follows a mixed approach that entails a dual application of strong and weak remedies. The guiding principle "*nemo ad facere compelle potest*" ("nobody can be forced to do [something]") serves as the aspirational criterion in the application of contract remedies. Article 1184, par. 2 of the French Civil Code allows the promisee to obtain specific relief, "when the original performance is still possible" (hence excluding its use in case of impossibility), but Article 1142 restricts the adoption of specific performance remedies when their use would infringe upon the individual freedom or autonomy of the promisor (Zweigert & Kotz, 1998, p. 475-9). In applying these principles, French-based civil law jurisdictions have come to distinguish between "obligations to give" (e.g., delivery of an existing good or transfer of title to land) and "obligations to do" (e.g., production of a not-yet-existing good or performance of a service), granting specific performance for the enforcement of obligations to give, and damages for the obligations to do (Treitel, 1988).[5]

(iii) *Weak-Remedy Systems.* The Anglo-American common law system stands at the other end of the spectrum, utilizing damages as the default remedy for breach of contractual

[5] The 1980 Vienna Convention on Contracts for the International Sale of Goods, the 2016 *Unidroit Principles for International Commercial Contracts*, and the 2020 *Principles of European Contract Law* adopt solutions reflecting the dual French approach (Zweigert & Kotz, 1998, p. 484-5; Lando & Beale, 2000).

obligations, with some equitable exceptions (Farnsworth, 1970; Hillman, 2019).[6] It should be noted, at this point, that U.S. courts quantify damages (the money compensation to be paid in case of breach) in a variety of ways. The most common measure is that of *expectation damages*—an amount of money equivalent to the value that the injured promisee was expecting to receive from the contract. Another measure of damages for breach of contract is *reliance damages*—an amount of money that restores the injured promisee's level of well-being before entering the contract. This measure is generally lower than expectation damages because the injured promisee does not recover the value she was expecting to gain from the contract—instead, reliance damages merely seek to make the promisee "whole."[7] A third measure of damages for breach of contract is *liquidated damages*—the amount of damages that the parties agreed upon at the time of the contract.[8] Liquidated damages clauses are commonly used to estimate compensation for losses that are difficult to measure precisely. Their prevalence varies considerably across industries.[9]

Given the variety of breach remedies employed by modern legal systems, one might question whether a specific legal remedy governing a contract genuinely influences the conduct of contracting parties. Comparative legal scholars have themselves questioned the practical relevance of the formal divide between common law's reliance on damages and civil law's emphasis on specific performance. Treitel (2003, p. 622) argues: "Although the theoretical frameworks diverge, in practice both systems often converge on similar outcomes, particularly once ancillary remedies and equitable discretion are taken into account."[10]

[6] See Restatement (Second) of Contracts § 359(1) (1981) ("Specific performance or an injunction will not be ordered if damages would be adequate to protect the expectation interest of the injured party."); U.C.C. § 2-716(1) (Am. L. Inst. & Unif. L. Comm'n 2022) ("Specific performance may be decreed where the goods are unique or in other proper circumstances."); see also Campbell Soup Co. v. Wentz, 172 F.2d 80, 82–84 (3d Cir. 1948) (granting specific performance for a contract to supply unique goods unavailable on the open market). Legal historians have traced the peculiar trajectories involved in the evolution of contract remedies. In the early common law, contract enforcement relied almost exclusively on specific performance but later progressed to an almost exclusive reliance on damages (Durfee, 1935 and Dawson, 1959). Civil law systems evolved in the opposite direction—from an almost exclusive reliance on money damages under classical Roman law to the inclusion of specific performance remedies in modern codifications (Dawson, 1959, Zweigert & Kotz, 1998).

[7] This measure of damages generally includes restitution of the price (if any price had been paid already), plus any additional contract-specific expenditures that the promisee made in reliance of the contractual promise.

[8] For a good—and pleasantly readable—introduction to U.S. contract law, see Hillman (2019). Chapter 5 provides a comprehensive treatment of damages remedies.

[9] In construction contracts, liquidation damage clauses are nearly universal, as delays often produce hard-to-quantify losses; some jurisdictions even mandate them. For example, until 2007, the University of California required liquidated damages clauses in all construction contracts over $10,000 (University of California Office of the President, 2007). Similarly, real estate agreements frequently include liquidated damages clauses; in Texas, they appear regularly in residential, commercial, farm and ranch, and unimproved property sales contracts (Adams, 2022). Conversely, liquidated damages provisions are less frequent in other commercial contexts, such as intellectual-property, service, and distribution agreements (Lexology, 2020).

[10] Empirical evidence supports this skepticism. Prior to 1970, Israeli contract law mirrored common-law practice, awarding damages as the primary remedy and reserving specific performance only as an exceptional equitable

Our experimental study seeks to provide further insight into this question. We look at the effects of damages vs. specific performance on how situations arising after contract formation are resolved. Specifically, we investigate whether, in the wake of a market shock, a legal system's remedy for unkept promises affects the contracting parties' willingness and ability to renegotiate a contract to allow a breach to occur (when a specific performance remedy is used) or to obtain an efficient performance (when a damages remedy is used). Understanding how parties react to alternative legal remedies presents a critical factor for consideration in contract design.

The theoretical backdrop of our experiment is Coase's (1960) theorem—having served as a starting point for several important contributions to the law and economics of remedies, including those by Calabresi and Melamed (1972), Schwartz (1979), Polinsky (1980), Cooter (1982), Ayres and Talley (1995), and Kaplow and Shavell (1996). When applied to breach of contract situations, the Coase theorem basically states that, in the absence of transaction costs, the choice of legal remedies is irrelevant to overall welfare. Most specifically, Kaplow and Shavell (1996) argued that property and liability rules would yield identical outcomes regarding ex-post efficiency when transaction costs are zero, as parties could costlessly bargain to an efficient allocation, providing the natural foundation for our rational choice hypothesis—if an exogenous shock changes the parties' respective costs and benefits under the original agreement, parties will bargain and reach an efficient agreement—with a resulting breach or performance of the contract, regardless of the legal remedy being used. In our incentivized experiment, we test the irrelevance of remedies proposition and confirm the theoretical prediction.

Our interest in testing the irrelevance proposition is motivated by several experimental studies of the Coase theorem, which have identified deviations from its theoretical predictions. Hoffman and Spitzer (1982, 1986), for instance, conducted influential experimental studies

relief for real property and unique goods. In 1970, Israel aligned its remedies with European civil-law systems by making specific performance the default remedy, subject to narrowly drawn exceptions. *Contracts (Remedies for Breach of Contract) Law*, 5730–1970 (Isr.) (1970). Empirical evidence, however, shows that this formal change did not translate into greater use:) find that the rate of specific performance awards actually fell after the reform (Anidjar, 2017; and Zamir et al., 2020). This suggests that parties' behavior is driven more by economic incentives than by the nominal default rule. Likewise, the Commentary to the PECL observes that the purported gap between common law and civil law remedies may be "more apparent than real" (PECL, Art. 9:101, Comment 3). The introduction of Article 1195 in the 2016 reform of the French Civil Code established a hardship clause that allows parties to request renegotiation of a contract if unforeseen circumstances make performance excessively onerous. This provision marks a significant shift in French contract law, as it explicitly favors renegotiation over forced performance and aligns French law more closely with international principles, thereby blurring the traditional distinction between civil law and common law approaches to contract adaptation (Cartwright & Whittaker, 2017; Fauvarque-Cosson, 2017; Rowan, 2017). Together, these voices caution against overstating the doctrinal divide.

examining bargaining behavior under conditions designed to test the theorem's assumptions. Contrary to the Coase theorem's expectation of invariant outcomes regardless of initial rights allocations, they found that participants' adherence to moral norms and fairness principles significantly influenced bargaining outcomes, leading to systematic deviations from efficiency and invariance predictions. See also the evidence of behavioral factors such as inequity aversion (Fehr & Schmidt, 1999; Bolton & Ockenfels, 2000; Charness & Rabin, 2002), which may lead to departures from Coase's theoretical prediction. Further, Aivazian et al. (2009) showed that when the bargaining "core"—the set of stable cooperative solutions—is empty, outcomes tend toward inefficiency, countering Coasean expectations. Collectively, these studies highlight how behavioral biases and institutional constraints substantially limit the Coase theorem's predictive accuracy.

Similarly, Kahneman et al. (1990, 1991) provided robust evidence of the endowment effect, showing that individuals' willingness to accept compensation (WTA) consistently exceeded their willingness to pay (WTP) in bargaining scenarios. This behavioral bias undermines the Coase theorem's assumption of symmetric valuations, thereby challenging its conclusion regarding allocation invariance.[11] In an executory contract setting, the sense of endowment may diminish with weaker remedies, such as damages, since buyers are aware that the promisor can avoid performance by paying compensation. As a result, buyers may become less attached to the good under a damage remedy due to the increased uncertainty of obtaining actual performance. Different remedies—damages vs. specific performance—may also induce a different framing of the entitlement, establishing different reference points, which can influence valuation via loss aversion or status-quo bias. These frames may also interact with inequity aversion, shaping parties' perceptions of what constitutes a fair or equitable outcome in bargaining.

Building on these considerations, we hypothesize two main behavioral channels through which remedies could matter. First, the *endowment effect* may be more pronounced under Specific Performance: the legal framing of a right to performance may strengthen the buyer's psychological sense of ownership, making them less willing to renegotiate. Second, *aversion to advantageous inequality* may make buyers more willing to leave a larger share of the surplus to the seller during renegotiation. However, since this preference symmetrically affects both treatments, we do not anticipate it producing systematic differences between

[11] See section 5.1.3 in Medema, 2020, for an extended discussion.

remedies. Hence, if a remedy effect were to appear, it would primarily operate through the first channel, i.e., a lower willingness to renegotiate by buyers under Specific Performance.

According to these behavioral intuitions, even when transaction costs do not pose an impediment to parties' renegotiation, legal remedies may affect the contracting parties' willingness and ability to renegotiate a contract in the face of exogenous shocks. These questions have thus far been left open by the extant empirical and experimental literature. Our experiment investigates whether the choice of alternative legal remedies affects contracting parties' renegotiation and division of the renegotiation surplus.

1.2 Literature Review

Breach remedies play a crucial role in contract law, especially in situations where writing complete contracts is not feasible due, for example, to information asymmetry and uncertainty about possible changes in the contractual circumstances (Rogerson, 1992). Among other purposes, legal remedies to breach of contract serve to protect parties that make contract-specific investments, to mitigate hold-up problems, and to foster efficient renegotiation of contractual terms (Williamson, 1985; Rogerson, 1992; Nöldeke & Schmidt, 1995; Coase, 2006).

On the choice of remedies for breach of contract, legal scholars have taken quite diverse positions. As recently pointed out by Parisi et al. (2024), both the consequentialist (economic) and the deontological (moralist) viewpoints consider the failure to perform on a promise excusable in at least some subset of cases, but their perspectives do not always converge on when contract breaches should be permitted with the payment of damages and when specific performance remedies should instead be made available to the non-breaching promisee (Birmingham, 1969; Barton, 1972; Warkol, 1998). At one end of the spectrum, standard economic analysis holds that if the promisor's gain from breach exceeds the promisee's loss, then allowing breach with payment of damages is socially desirable, as it increases joint welfare relative to performance (Posner, 1999, 2009; Shavell, 2006, 2009). At the other end of the spectrum, deontological philosophers of contract law consider the moral duty to keep one's promises to constitute a foundational principle of contracts, which should not be brushed aside based on cost-benefit analyses (Sidhu, 2006; Mather, 1999; Fried, 1981; Shiffrin, 2009). Some law and economics scholars have similarly argued in favor of specific performance remedies, observing, among other concerns, that judicial liquidation of damages may be socially costly

(Kronman, 1978; Schwartz, 1979; Sloof et al., 2006; Depoorter & Tontrup, 2012).[12] As shown by Baron and Wilkinson-Ryan (2009) and Bigoni et al. (2017), the layperson's views about the excusableness of non-performance are surprisingly nuanced, embracing the consequentialist (economic) reasoning in some cases and the deontological (moralist) view in other cases.[13] Rachlinski and Jourden (1998) analyzed the well-known legal case *Boomer v. Atlantic Cement Co.*[14] and observed that, even after securing substantial damages, homeowners appealed to obtain an injunction—suggesting they valued injunctive relief more than monetary damages. The authors explain this finding suggesting that specific performance conveys a deeper sense of psychological ownership by preserving the rightholder's power to exclude others—an essential element missing when rights are enforceable only through damages. Departing from the Coasean account of remedies, the authors show that remedial rules also send normative signals, activating ownership psychology and shaping litigants' behavior in ways not predicted by traditional law and economics models.

In an earlier experimental study, Depoorter and Tontrup (2012) reached a similar conclusion, showing that the granting of a specific performance remedy signals a promisor's duty and triggers a stronger sense of obligation than damages alone: in the experiment, subjects faced with specific performance were significantly more likely to fulfill their promises—an "expressive effect" of the remedy.

Bigoni et al. (2017) extended the inquiry, allowing parties to renegotiate, to study whether the promisor's motive for the breach—avoidance of a loss vs. pursuit of more lucrative offers—affects the parties' willingness and ability to renegotiate. In that experiment, parties always carried out renegotiation under a default specific performance remedy and needed to negotiate out of the injunctive remedy to avoid performance. The experiment showed that the motive for the breach affected the parties' willingness to renegotiate and the splitting of the surplus from renegotiation.

---

[12] Scholars in contract law and economics argued that a damages remedy is an efficient default remedy for contract breach, since it induces breach only if the cost of performance for the promisor outweighs the value of performance for the promisee (Birmingham, 1969; Barton, 1972). Some other scholars argued in favour of specific performance as the most efficient default remedy, since it embraces the moral obligation that promises should be kept (Depoorter and Tontrup, 2012), and because damages may impose unnecessary costs (Kronman, 1978; Schwartz, 1979). However, no evidence has been presented to see how the views of ordinary contracting parties align with these opposing claims.

[13] The survey-based studies and economic experiments conducted by Baron and Wilkinson-Ryan (2009) and Bigoni et al. (2017) have shown that ordinary people have greater tolerance for contract breaches when the promisor seeks to avoid unanticipated losses (i.e., loss-avoiding breaches) but are less willing to excuse performance when the promisor breaches to pursue a profit (i.e., gain-seeking breaches).

[14] *Boomer v. Atl. Cement Co.*, 26 N.Y.2d 219, 309 N.Y.S.2d 312, 257 N.E.2d 870 (1970).

Engel and Freund (2017) explore whether the choice of remedy—expectation damages versus specific performance—affects parties' willingness to enter into mutually beneficial contracts ex ante, and therefore if it affects efficiency. The authors find that when specific performance is not guaranteed, trade rates fall significantly when valuation differences are large enough to make the gains from trade salient, even though expectation damages would fully compensate the non-breaching party. This result implies that, although expectation damages deliver ex post efficiency, they can discourage ex ante contract formation—contradicting the Coasean irrelevance of remedies proposition, as far as contract formation is concerned.

Relatedly, Sloof et al. (2003) compared reliance investments under different breach remedies. They found that specific performance led to higher levels of contract-specific investment, suggesting that stronger remedies increase perceived commitment and encourage greater ex ante investment. Sloof et al. (2006) extended this analysis by allowing for renegotiation following a breach. While they confirmed that specific performance induced higher investments, they also found that this did not necessarily lead to more efficient outcomes. Lewinsohn-Zamir (2013) finds that both laypeople and professionals tend to prefer in-kind remedies over monetary ones in hypothetical breach scenarios. Unlike designs that emphasize the impact of remedies on reliance investment or contract entry, our focus is on how remedy framing affects behavior during renegotiation itself.

Our experiment differs from these earlier contributions because we allow the default contract remedy to vary in order to analyze whether individuals' incentives to renegotiate and division of surplus differ when the default legal remedy varies. The experiment attempts to replicate situations where changes in the original conditions (at the time of contract formation) lead parties to reconsider their original contractual commitment. Such changes to conditions include changes in the original costs and the rise of new opportunities, which may alter the opportunity cost of the parties' original contractual engagement. We refer to these latter factors as the "outside options" faced by the contracting parties. Our analysis delivers novel findings regarding parties' renegotiation behavior in the shadow of alternative legal remedies. The results will provide valuable information to transactional lawyers in designing optimal contracts and to lawmakers and judges entrusted with making policy choices over which breach remedies to apply in the face of exogenous market shocks.

2. Experimental Design.

The experiment comprises two parts. In Part 1, we consider a modified version of the "Contract-Breach Game" introduced by Bigoni et al. (2017), where a market shock *may* occur after the parties have agreed to a binding contract. Once the shock occurs, the parties may renegotiate the contractual terms.[15] The key, novel aspect of our design is to allow the comparison of renegotiation outcomes under two *breach remedies*: a damages remedy (*Damages* treatment) and a specific performance remedy (*Specific Performance* treatment). This modified "Contract-Breach Game" à la Bigoni et al. (2017) allows us to study how frequently an efficient agreement *to breach* is reached under specific performance, and how frequently an efficient agreement *to perform* the contract is reached under damages. Furthermore, this game allows us to analyze whether the compensation paid by the promisor to the promisee differs between alternative breach remedies, and between alternative reasons to breach. In Part 2, all participants take part in a Dictator Game, to control for their inequality aversion.

Contract-Breach Game. In Part 1, participants are randomly assigned the role of buyer (B) or seller (S) and are matched in pairs. Production costs (C) are initially set to 40 and the value of the good (V) for the buyer is set to 60, yielding a total surplus (π_{Total}) of 20. Unlike in Bigoni et al. (2017), the parties are forced into a contract in which the price, P, is initially exogenously set to 50. This means that the party cannot refuse to enter the initial contract in Stage 0, where buyers and sellers are simply informed about the contract. This was intended to avoid any difference in entry rates due to different remedies. While this is a potentially interesting margin—and remedies might indeed affect the decision to enter a contract in the first place or influence the price—such differences could have complicated our subsequent analysis of surplus division.

Table 1 provides an overview of the payoffs for sellers and buyers in all relevant cases. A shock may occur, modifying the initial parameters: both production costs and the value of an outside option (for either the buyer or the seller) increase. When shocks occur, forcing specific performance (in the Specific Performance treatment) or allowing the breach of the

[15] It should be noted that under the U.S. "pre-existing duty rule," a price modification following a change in circumstances must be "fair and equitable" to be enforceable. See Restatement (Second) of Contracts § 73 (1981); U.C.C. § 2-209 (Am. L. Inst. & Unif. L. Comm'n 2022). Our experiment does not aim to test the validity, application, or desirability of this rule, and our experimental design allows parties full freedom in the range of renegotiation.

contract (in the Damages treatment) would yield an aggregate loss. In both cases a contract renegotiation would be efficient. Importantly, the total surplus from renegotiation is held constant across treatments.

Table 1: Overview of the stage game

Stage	Specific Performance	Damages
Stage 0 **Initial parameters**	Cost of production for seller: $C = 40$ Value of good for buyer: $V = 60$ Contract price: $P = 50$	
Stage 1 **Shock** **(prob 60%)**	$C' = 60$ $V' = 60$ Outside option for S: $O_S = 80$	$C' = 60$ $V' = 80$
Stage 2 **Enter renegotiation**	***B* can enforce performance,** and the contract is performed or move to stage 3. **Performance is inefficient.**	***S* can breach** and the contract is **not** performed or move to stage 3. **Breach is inefficient.**
In case of no renegotiation	The contract is performed: $\pi_S^d = 50-60 = -10$ $\pi_B^d = 60-50 = 10$ $\pi_{total}^d = 0$	Damages (D = 60) are paid: $\pi_S^d = 50-60 = -10$ $\pi_B^d = 60-50 = 10$ $\pi_{total}^d = 0$
Stage 3 **Renegotiation phase**	S and B renegotiate to allow a breach (buyer is compensated with a transfer t) $\pi_S^r = 80-60-t \in [-10, 10]$ $\pi_B^r = t \in [10, 30]$ $\pi_{total}^r = 20$ $10 \leq t \leq 30$	S and B renegotiate to fulfil the contract at a new price $P' > P$ $\pi_S^r = P' - 60 \in [-10, 10]$ $\pi_B^r = 80 - P' \in [10, 30]$ $\pi_{total}^r = 20$ $50 \leq P' \leq 70$

A key feature of our design is that the economic consequences of failing to renegotiate are held constant across treatments. In both the Damages and the Specific Performance conditions, if the parties do not reach an agreement after a shock, the seller incurs a loss of 10 tokens, the buyer earns 10 tokens, and total surplus is zero. What differs between treatments is only the entitlement structure—that is, which party has the right to enforce the inefficient outcome. Under Specific Performance, the buyer holds the right to enforce the original contract, so the seller must make a compensatory offer to persuade the buyer to cancel it. Under Damages, by contrast, the seller has the right to breach upon payment of damages, so the buyer must make a counter-offer to induce the seller to perform. In this way, the two treatments generate mirror-image bargaining situations: the outside options and the size of the renegotiation surplus are identical, while the identity of the party who must make concessions is reversed. In the following, we describe all the stages of the game in detail (see Table 1 for an overview).

Once randomly matched in pairs, buyers and sellers go through the following steps.

Stage 0: The buyer and the seller are informed about the production costs and the benefits from the transaction.

Stage 1: An exogenous shock may happen with 60% probability and lead to an increase in production costs, $C' = 60$, in both treatments.[16] At the same time, in the Specific Performance treatment, the seller has an opportunity to sell the good to a different buyer for a price of 80, while in the Damages treatment, the value of the good for the buyer increases to 80.

Stage 2: The parties must decide whether to renegotiate the contract. Given the increased production costs, the seller has an interest in breaching the contract. Renegotiation starts if both parties decide to enter the renegotiation stage.

- Under a Specific Performance remedy, the buyer can force the performance of the contract, and the seller cannot unilaterally choose to breach. If the seller chooses to perform his existing obligation, this period is over. By performing, the seller suffers a loss, while the earnings of the buyer remain unchanged compared to stage 0. The total surplus is equal to zero. Alternatively, the seller can seek cancellation of the contract by offering compensation *t* to the buyer.
- Under a Damages remedy, the seller may unilaterally choose to breach the contract by paying liquidated damages.[17] To create conditions where parties have equal incentives to renegotiate—whether to secure efficient performance or to allow an efficient breach—we tied liquidated damages to the value of performance assessed at the time the contract was

[16] We considered it important to include both shock and no-shock periods in the design for two main reasons. First, a shock should be perceived as a possible, but not certain, state of the world. To reinforce this notion, the state of the world was randomly determined in each period through a random draw. Second, periods without a shock serve as a benchmark to assess participants' understanding of the stylized environment proposed in the experiment, as renegotiation is not efficient in those cases. The selected probability of a shock (60%) aims to balance the need for uncertainty regarding shocks with the desire to obtain repeated observations in periods where the initial conditions change.

[17] Parties have incentives to renegotiate whenever contractual damages are undercompensatory—that is, when the buyer's actual loss exceeds the damages recoverable in the event of breach. For simplicity, we illustrate this scenario by setting liquidated damages at the expectancy value at the time the contract is formed. Comparable incentives for renegotiation would arise under other undercompensatory remedies, such as reliance damages or restitution.

formed, $D = V$.[18] If the contract is breached, this period is over. By breaching and paying damages, the seller suffers a loss, while earnings for the buyer remain unchanged compared to stage 0. The total surplus is equal to zero. Alternatively, the parties can renegotiate the contract and agree to perform for a new price *P'*.

Stage 3: If the parties decide to renegotiate, they have 60 seconds to reach an agreement on the compensation *t* payable to the buyer (Specific Performance treatment) or the new price *P'* (Damages treatment). We opted for a continuous bargaining process for both theoretical reasons and to enhance the realism of bargaining situations, in which parties typically do not follow a strict sequence in making (counter)offers. The time limit was set—based on prior studies—to allow for meaningful negotiation while avoiding excessively long waiting times and sessions.[19] Specifically, under each of the two remedies, renegotiation takes place as follows.

- Under Specific Performance, the seller can offer any positive integer *t* between 10 and 30 to compensate the buyer for the cancellation of the contract. The buyer can either accept the offer or demand a higher amount *t* (always between 10 and 30), and the seller can accept or reject his demand. Offers and demands can be adjusted at any time, during the 60-second renegotiation stage. If an offer or demand is agreed upon, the original contract is resolved, and the parties' payoffs are determined according to the agreed terms. If no agreement is reached by the end of the renegotiation stage, the original contract is enforced through specific performance.
- Under Damages, the buyer can offer any price *P'* between 50 and 70 to convince the seller to carry out the performance instead of breaching the contract. The seller can either accept the offer or propose a higher (between 50 and 70) price. Offers and proposals can be revised at any time, during the 60-second renegotiation stage. If an offer or proposal is accepted, the contract is resolved, and the parties' payoffs are determined according to the

[18] Under the principle of expectation damages in contract law, damages would be tied to the buyer's benefit of performance assessed at the time of breach. Friedmann (1989) argues that expectation damages often fail to compensate for non-monetary losses (such as reliance investments, reputational damage, and the moral force of promise-keeping), meaning that a breach that appears "efficient" under a simple calculus can leave promisees worse off in reality. Along these lines, Goetz and Scott (1977) from an ex ante perspective, parties optimally use liquidated damages clauses to allocate risk and reduce risk of undercompensation, minimizing the need for post breach litigation.

[19] While the 60-second time limit may seem stringent, it is important to note that there were only two possible states of the world (shock or no shock), all relevant parameters were known from the outset, and the stage game was repeated multiple times. The high rate of successful renegotiations (see Section 4.1) suggests that participants did not feel unduly pressured or rushed during this stage.

agreed terms. If no agreement is reached by the end of the renegotiation stage, the contract is breached, and the seller will pay damages, compensating the buyer for the expected value of the performance, $D = V$.

In our experiment, the game was repeated for 25 periods and roles were fixed throughout the experiment. The game was repeated for at least three reasons: (i) to provide participants with learning opportunities; (ii) to reduce measurement error; and (iii) to exploit heterogeneity at both the group and individual levels in order to examine the role of other-regarding preferences. At the beginning of each period, buyers and sellers were paired randomly, within matching groups of 6 participants (3 buyers and 3 sellers).[20] Their identities remained anonymous throughout the experiment to avoid any reputational effect, while preserving the one-shot nature of the Contract-Breach Game. At the end of each round, the parties were informed about the outcome of the renegotiation phase, their own earnings, and the earnings of their counterpart. Subjects could always see their cumulative earnings (including their initial endowment) on the screen. Shocks occurred with 60% probability and the random draw was performed at the matching-group level. We predetermined the sequence of shocks over the 25 periods for six sessions, and each predetermined sequence was used once for the Damages treatment and once for the Specific performance treatment. This generated a balanced sequence of shocks across the two treatments.

After the Contract-Breach Game, participants played a dictator game in which they were asked to split 20 tokens between themselves and an anonymous counterpart. This task serves as a proxy for inequality aversion, which may interact with the shift in the sense of entitlement possibly determined by the alternative remedies, and influence how the surplus from renegotiation is distributed, in the two treatments. Finally, participants completed a short questionnaire on individual characteristics (gender, education level, field of study, etc.). The instructions for both treatments and the Dictator Game are provided in the Appendix, together with the final questionnaire.

Procedures. A total of 288 subjects, equally divided across treatments, participated in the experiment. Subjects were recruited via ORSEE (Greiner, 2015), and all 12 sessions were run in English at the BLESS lab in Italy. The average age of our subjects was 24.7, 53.12% identified as females and 39.6% as males, and 92.4% were born in Italy. These characteristics

[20] We determined the size of each matching group by taking into account logistical constraints (e.g., lab seating), statistical power considerations, and the need to ensure sufficient variation in matched partners.

are balanced across treatments (see Table A1 in the Appendix). The experiment was programmed and conducted using the z-Tree software (Fischbacher, 2007). Each session lasted about 90 minutes, and the average payment was 12.9 euros. The study was approved by the ethical committee of the University of Bologna (Protocol N. 0156345) and was pre-registered (AEARCTR-0013798: https://doi.org/10.1257/rct.13798).

Outcomes of interest and testable hypotheses. In our set-up, parties have perfect information on each other's payoff functions, there are no transaction costs, and contract remedies are enforced instantly and without cost. Within this framework, the Coase theorem predicts that the parties will always reach a surplus-maximizing agreement, regardless of the remedies in place. Since the payoffs parties obtain when bargaining fails are also identical across treatments, we expect the difference between remedies to have no impact on how the surplus is distributed between them. However, different models offer varying predictions on how the parties will resolve the bargaining problem. According to the Nash bargaining solution (Nash, 1950) the parties should agree on an outcome that yields zero surplus to the seller (i.e. $\pi_S^r = 0$ and $\pi_B^r = 20$), since this is the solution that maximizes the product of their excess utilities, given by:

$$(\pi_S^r - \pi_S^d)(\pi_B^r - \pi_B^d)$$

The same prediction is also supported by alternative bargaining solutions (Kalai & Smorodinsky, 1975; Kalai, 1977), which in our parameterization lead to the same outcome.[21]

However, behavioral phenomena such as the endowment effect, loss aversion, and diminishing sensitivity may lead to deviations from this prediction and to differences across treatments. Although our experimental design holds constant both the strategic structure and material payoffs across treatments, it introduces a salient framing contrast during the bargaining stage. Under Specific Performance, the seller—facing a loss of 10 tokens if renegotiation fails—must compensate the buyer to cancel the contract. Under Damages, by contrast, the buyer must offer the seller a higher price to induce performance. These differing framings may establish distinct reference points for participants, leading to asymmetric

[21] In general, the Kalai–Smorodinsky (1975) solution and Kalai's proportional solution (1977) are distinct. In our design, however, the feasible payoffs after renegotiation range from (–10, 30) to (10, 10), so the gains from the disagreement point are 20 tokens for both parties. This symmetry causes the Kalai–Smorodinsky condition to collapse to equal absolute gains, making the two solutions coincide and yielding the same outcome as the Nash solution (0 for the seller, 20 for the buyer).

perceptions of gains and losses. As a result, the framing could influence both the likelihood of agreement and the division of surplus. In this way, the legal remedy may shape how parties perceive their entitlements, even when objective payoffs remain unchanged.

In light of these considerations, we focus on three main outcomes of interest, and three main null hypotheses.

- **Renegotiation entry:** Percentage of contracting parties that enter the renegotiation phase after a shock—these are cases where renegotiating the contract would always be efficient.
  HP1: Different remedies do not affect the parties' ability to enter an efficient contract renegotiation.
- **Renegotiation success:** Percentage of contracting parties that successfully reach an agreement, conditional on entering the renegotiation stage.
  HP2: Different remedies do not affect the probability that an agreement is reached, conditional on entering renegotiation.
- **Surplus to seller:** Percentage of the renegotiation surplus captured by the seller. The variable is defined only for those contracting parties who managed to successfully renegotiate the contract after a shock.
  HP3: For all the contracting parties who renegotiated the contract, the division of the aggregate surplus is not affected by the choice of remedy.

It is important to note that our parameters were calibrated to make renegotiation relatively easy, following the example of Bigoni et al. (2017). Our main focus is on the surplus gained by the seller, and to this end, we aimed to maximize the number of available observations while avoiding excessive burden on subjects through too many repetitions of the stage game. This, of course, comes at the cost of possibly failing to detect some effects on the renegotiation success rate, as we might be operating close to a ceiling.

3. Empirical approach

Following the pre-registration, we test each of our three main hypotheses by means of a two-pronged approach. The first approach is more conservative and is based on non-parametric two-tailed Wilcoxon Mann-Whitney tests, which test the difference in the distribution of the outcomes of interest between treatments. The unit of observation in this case is the average outcome per matching group, across all rounds, yielding 48 independent observations in total. This allows us to detect an effect of 0.85 of a s.d. with $\alpha = 0.05$ and $(1 - \beta) = 0.8$. Based on the data collected by Bigoni et al. (2017), this corresponds to a difference of 3 percentage points in the percentage of contracting parties who enter renegotiation when an exogenous shock takes place (*HP1*), a difference of 5 percentage points in the rate of success of renegotiation (*HP2*), and a difference of 0.9 units (out of 20) in the surplus allocated to the Seller in the renegotiation stage, when renegotiation is successful.[22]

The second approach relies on the panel structure of our dataset and allows us to control for learning dynamics across periods. For each of the three outcomes mentioned in Section 2, we ran a linear regression based on the following equations:

$$Y_{i,t} = \beta_0 + \beta_1 D_i + \beta_2 t + \eta_i + \epsilon_{i,t} \quad (1)$$

$$Y_{i,t} = \beta_0 + \beta_1 D_i + \beta_2 t + \beta_3 D_i t + \eta_i + \epsilon_{i,t} \quad (2)$$

where $Y_{i,t}$ is the outcome of interest, in matching group i and period t, D_i is a dummy variable taking value 1 if matching group is assigned to the *Damages* treatment and 0 if it is assigned to the *Specific Performance* treatment, and η_i is the matching-group specific error component. Since we only focus on the periods in which a shock takes place, with this specification we will have an unbalanced panel of 48 individual matching groups and approximately 15 periods. Based on the simulations obtained via the Stata pc_simulate package (Burlig et al., 2020), we detected a variation of 4 percentage points in the share of contracting parties who enter renegotiation when an exogenous shock takes place (*HP1*), a difference of 6 percentage points in the rate of success of renegotiation (*HP2*), and a difference of 0.9 units (out of 20) in the surplus allocated to the Seller in the renegotiation stage, when renegotiation is successful, with $\alpha = 0.05$ and $(1 - \beta) = 0.8$.

Heterogeneity. Other-regarding preferences have been shown to play an important role in bargaining. For instance, in the context of the Contract-Breach Game, Bigoni et al. (2017) showed that inequality-averse subjects accept low offers more often in cases of loss-avoiding breaches than gain-seeking breaches. In light of this evidence, we conducted a heterogeneity analysis leveraging data from the Dictator Game played in Part 2 of the experiment.

We classify subjects into 'inequality-tolerant' and 'inequality-averse' based on a median split of the allocation choice in the Dictator Game. For each matching group, we construct the dummy variable *Altruistic Buyers* (B_i), taking the value 1 if the majority of the

[22] The power calculation was performed using G*Power3 (Faul et al., 2007; 2009).

buyers in the group are inequality-averse. Similarly, for each matching group, we will construct the dummy variable *Altruistic Group* (G_i) taking the value 1 if the majority of the buyers and sellers in the group are inequality-averse. We ran a linear regression based on the following equation:

$$Y_{i,t} = \beta_0 + \beta_1 D_i + \beta_2 B_i + \beta_3 D_i B_i + \beta_4 t + \eta_i + \epsilon_{i,t} \quad (3)$$

where $Y_{i,t}$ is each of the three outcomes of interest described in Section 2, in matching group i and period t, D_i is a dummy variable taking the value 1 if the matching group is assigned to the *Damages* treatment and 0 if it is assigned to the *Specific Performance* treatment, and η_i is the matching-group specific error component. We repeated the same analysis in equation (3) for G_i.

We also looked at the periods in which there was no initial shock to check if other-regarding preferences might have played a role. In these periods, one should expect no renegotiation, but an inequality-averse buyer could be prompted to use renegotiation to reduce the unbalance between buyers' and sellers' earnings from the bargaining phase. A buyer's willingness to renegotiate in periods without shocks could create an opportunity to offer the seller a more favorable price.

4. Results

We have a total of 1200 observations at the matching group level, 680 (56.7%) of which involved a shock.[23] In the periods without shocks, only 16.0% of the contracting parties entered the renegotiation stage and only 27.7% of them successfully reached an agreement (i.e., in the absence of a shock, only 4.4% of the agreements were modified). When entering renegotiation without a shock, the parties confirm their willingness to enter the contract, and simply renegotiate on the selling price. As one might expect, when there are no changes to the initial conditions the total surplus is divided almost equally among the parties. The surplus to the seller in the Specific Performance treatment was 57.5% and in the Damages treatment was 45.8%, which is very close to the equal split they would have obtained from the original agreement, without renegotiation.

[23] The sequence and the number of shocks over the 25 periods varied across sessions and matching-groups; the minimum (maximum) number of shocks observed by a matching group was 6 (19).

In the remainder of the paper, we focus only on the periods in which a shock occurred and test for treatment differences in our three main variables of interest: renegotiation entry, renegotiation success, and surplus to the seller (see Table 2).[24]

4.1 Renegotiation results for periods with a shock

Overall, 94% of the contracting parties entered the renegotiation stage when a shock to the original conditions occurred (specifically, 93.1% renegotiated under a Damages remedy and 94.9% under a Specific Performance remedy).[25] We fail to find any statistically significant difference in the renegotiation entry across treatments (Mann–Whitney U test, 24 obs. each treatment, $p = 0.796$, two-sided). Models 1 and 2 in Table 3 provide evidence from regressions exploiting the panel nature of the data where the dependent variable is Renegotiation Entry (%). Regression results confirm that no significant treatment difference emerges, but they also reveal that, as one might expect, there is some learning, with matching-groups becoming more likely to enter the renegotiation as the game progresses (see also Figure 1, panel (a)). Yet, there is no difference in this learning process across treatments. This finding is not surprising considering that parties are not forced in any way to reach an agreement that they do not consider satisfactory.

Table 2: Descriptive statistics

	Renegotiation Entry	Renegotiation Success	Surplus to Sellers
Damages	0.931	0.864	0.155
(s.d.)	*(0.080)*	*(0.062)*	*(0.089)*
c.i.	[0.897-0.964]	[0.837-0.890]	[0.118-0.192]
Specific Performance	0.949	0.839	0.178
(s.d.)	*(0.048)*	*(0.069)*	*(0.077)*
c.i.	[0.928-0.969]	[0.809-0.868]	[0.146-0.211]
Damages vs. Specific Performance (p-val.)	0.796	0.180	0.257

Notes: All p-values are calculated using a two-sided Mann–Whitney U test.

We now focus on the contracting parties who entered the renegotiation stage after a shock and consider the share of contracting parties who managed to successfully reach an agreement. Overall, 85% (86% and 84% in Damages and in Specific Performance, respectively) of renegotiating parties reached an agreement. This is quite remarkable given the short amount of time available to renegotiate and shows that our subjects were especially aware

[24] All the analyses presented in this section were pre-registered.
[25] Both parties in the pair need to agree to enter the renegotiation. If we look at individual-level data, out of 4080 decisions only 117 did not agree to enter the renegotiation stage.

of the efficiency gains obtainable through renegotiation. We fail to find any statistically significant difference in the renegotiation entry across treatments (Mann–Whitney U test, 24 observations in each treatment, $p = 0.180$, two-sided). It is worth noting that we may fail to find a difference between the two treatments, as—by construction—the parameters were designed to favor renegotiation, and we may be close to a ceiling. However, the confidence intervals clearly show that the renegotiation rates are significantly lower than 100%, so in both treatments there would be scope for further improvement. Models 3 and 4 in Table 3 report results from panel regressions and confirm that the rate of successful renegotiation does not depend on the treatment. Interestingly, we do not find any strong sign of learning over time as the coefficient of the variable period is zero (as confirmed also by Figure 1, panel (b)).

Table 3: Regression Analysis

	Renegotiation Entry (1)	Renegotiation Entry (2)	Renegotiation Success (3)	Renegotiation Success (4)	Surplus to Sellers (5)	Surplus to Sellers (6)
Damages *(d)*	-0.014	-0.013	0.023	0.002	-0.024	0.003
	(0.019)	(0.026)	(0.020)	(0.037)	(0.024)	(0.029)
Period	0.005***	0.005***	0.001	-0.000	-0.005***	-0.004***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Damages *(d)* x Period		-0.000		0.002		-0.002*
		(0.001)		(0.002)		(0.001)
Intercept	0.879***	0.879***	0.830***	0.842***	0.244***	0.231***
	(0.016)	(0.018)	(0.021)	(0.026)	(0.019)	(0.020)
R-squared	0.069	0.069	0.003	0.004	0.069	0.072
N. obs.	680	680	680	680	670	670

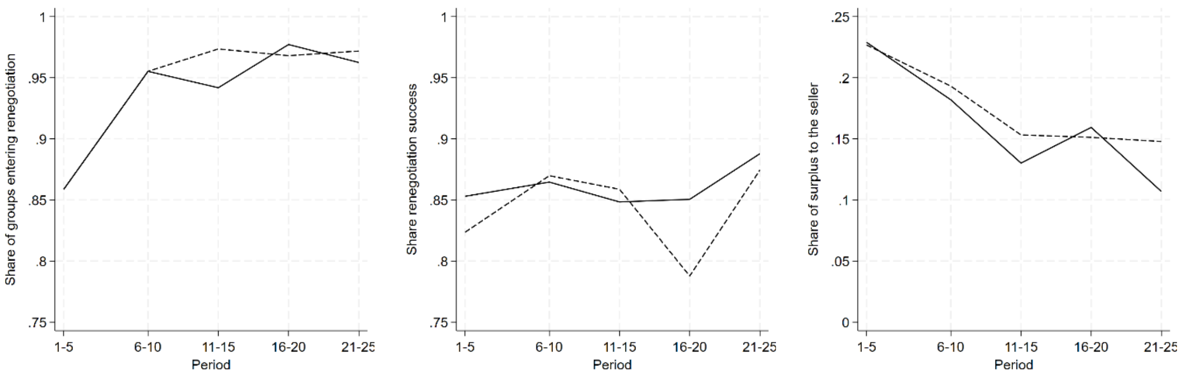
Notes: GLS panel-data regression with random effect. The symbol *** indicates significance at the 1%. Dummy variables are denoted with *(d)*.

So far, we have established that in both treatments the vast majority of our experimental participants choose to renegotiate and are able to reach an agreement after a shock. We now study how the surplus is divided between buyers and sellers. In particular, we consider the share of the surplus earned by sellers who possess less bargaining power relative to buyers, as these sellers suffer a loss when disagreement occurs. As a result of the renegotiation, sellers earn only a small share of the available surplus: 15% in the Damages treatment and 18% in the Specific Performance treatment, revealing no statistically significant effect of remedies on the division of the surplus (Mann–Whitney U test, 24 obs. each treatment, $p = 0.257$, two-sided). Panel (c) in Figure 1 shows the evolution over time, highlighting that the share allocated to sellers declines in both treatments. This suggests that buyers learn to better exploit their bargaining power, leaving a smaller share of the surplus to sellers. Models 5 and 6 in Table 3 report results from panel regressions formally testing for treatment differences and learning effects. While we do not find any clear difference across treatments, the variable Period is negative and significant, confirming the downward trend consistent with buyer learning. Yet, it is quite puzzling to notice that the effect of the learning process slightly diverges between the two treatments (*Damages (d) x Period*) with sellers' earnings declining more steeply over time in Damages rather than in Specific Performance (see also Figure 1, panel (c)).

Our findings so far are consistent with the prediction of Coase's (1960) theorem, according to which, in the absence of transaction costs, the choice of legal remedies is irrelevant to overall welfare. Within the precision of our estimates, we find no evidence that the choice of remedy affects the parties' willingness to renegotiate, their ability to reach agreement, or the division of the surplus. While these results should not be interpreted as a formal confirmation of the null hypothesis, the observed patterns suggest that any treatment effects, if present, are likely to be small. Of course, our evidence comes from a highly controlled environment where parties are able to enforce remedies costlessly (in both treatments specific performance and damages can be enforced costlessly), buyers have not made reliance investments, and the parties cannot hedge against the risk of a shock.

Figure 1: Behavior over time

a) Renegotiation entry b) Renegotiation success c) Surplus to sellers



Notes: Solid lines denote the Damages treatment; dashed lines denote Specific Performance.

In the remainder of the analysis, we conduct a more in-depth examination of the influence of inequality aversion on the observed results.

4.2 Results for the heterogeneity analysis

To better understand the renegotiation dynamics, we used answers to the Dictator Game played in Part 2 of the experiment. Out of 20 tokens, participants kept for themselves on average 14.5 tokens (median: 13 tokens), with roughly one third of the participants keeping everything for themselves and one third sharing equally. The average number of tokens kept was similar across both roles (i.e., buyers vs. sellers) and treatments (Damages vs. Specific Performance). The distribution of allocations in our Dictator Game is comparable to Engel's (2011) meta-study. About one third of subjects gave nothing, and a significant fraction gave close to half. While Engel reports a somewhat more diffuse pattern (with 16–17% giving half), our participants exhibit a more polarized distribution, with 34% giving nothing and 33% giving half. This pattern falls well within the heterogeneity reported in prior dictator game experiments.

Following our pre-analysis plan, we classified buyers and sellers as "inequality-averse" if they kept for themselves a smaller or equal number of tokens relative to the median in the population. We opted for a median split to avoid any arbitrary cut-off value, in light of the fact that giving in the Dictator Game can be interpreted in different ways (inequality aversion, altruism, or other fairness concerns), and that in a bargaining experiment what is relevant are the relative positions of the participants. Hence, our measure of inequality aversion must be interpreted in relative terms and with reference to our participant pool, and not as an absolute aversion based on some general parameters. We then defined two dummy variables:

- *Altruistic Buyers*[26] taking the value 1 if the majority of the buyers in the matching group are inequality-averse
- *Altruistic Groups* taking the value 1 if the majority of the buyers and sellers in the matching group are inequality-averse

Overall, 46% of the matching groups are classified as *Altruistic Buyers*, and 31% as *Altruistic Groups.*[27] In both cases, the term *altruistic*—used to refer to buyers or groups with greater aversion to advantageous inequality—should be interpreted in relative terms, as the goal is

[26] Although we refer to these groups as *Altruistic Buyers* and *Altruistic Groups* in the tables and text, this terminology serves only as a label for participants with higher measured aversion to advantageous inequality.

[27] 50% of matching groups in Damages and 41.7% of matching groups in Specific Performance are classified as *Altruistic Buyers*. 37.5% of matching groups in Damages and 25% of matching groups in Specific Performance are classified as *Altruistic Groups.* For both *Altruistic Buyers* and *Altruistic Groups*, there is no significant difference across treatments (Mann–Whitney U test, p-value > 0.2, two-sided).

simply to create a proxy for some characteristics possibly relevant for the bargaining phase. Our conjecture is that, in a context without reputation formation or partner selection, players with stronger aversion to advantageous inequality are more likely to make better offers and accept lower ones— as they care more about reducing payoff differences.[28]

We first focus on buyers, as they possess relatively stronger bargaining power and thus are more likely to shape the bargaining process. This stronger bargaining power arises because, in the event of failed renegotiation, the buyer's disagreement payoff is +10 tokens, while the seller's is –10 tokens; this asymmetry gives buyers a more favorable outside option and thus greater leverage during bargaining. Table 4 reports descriptive statistics for our three outcome variables of interest, for two categories of matching groups. The first category includes groups with at most one inequality-averse buyer (i.e., *Selfish-Buyer* groups), the second category comprises groups with two or three inequality-averse buyers (i.e., *Altruistic-Buyer* groups). The composition of the group only marginally affects the likelihood of entering the renegotiation, which is extremely high for both types of groups, though the likelihood is slightly but significantly lower for *Altruistic-Buyer* groups under Specific Performance. Similarly, buyers' inequality aversion does not seem to affect the probability of successful renegotiation. Interestingly, some differences arise with respect to the division of surplus. Under both remedies, Altruistic-Buyer groups tend to leave a larger share of the surplus to sellers (see last two columns of Table 4). However, the difference is small in magnitude and statistically significant only under Damages. This adds to previous evidence on the role of individual preferences on bargaining outcomes. If we compare the two treatments, we see that when buyers are less inequality averse, they end up reaching better deals for themselves (i.e., they leave a smaller share of the surplus to the sellers) under Damages. Less inequality-averse buyers have more leverage in their renegotiation under a Damages remedy.[29]

[28] While we adopt inequality aversion as our main interpretive framework, some scholars interpret Dictator Game giving as reflecting generosity or altruism more broadly. In our setting, however, both interpretations yield similar behavioral predictions.

[29] These results are confirmed by panel regressions presented in Table A.2 in Appendix. Results are qualitatively the same, if we consider *Altruistic* and *Selfish Groups* rather than *Altruistic* and *Selfish Buyer.* Results are available upon request to the authors.

Table 4: Descriptive statistics on the role of altruistic buyers

	Renegotiation Entry **Selfish**	Renegotiation Entry **Altruistic**	Renegotiation Success **Selfish**	Renegotiation Success **Altruistic**	Surplus Sellers **Selfish**	Surplus to Sellers **Altruistic**
Damages	0.952	0.910	0.849	0.878	**0.120**	**0.190**
(s.d.)	*(0.048)*	*(0.100)*	*(0.068)*	*(0.056)*	*(0.068)*	*(0.096)*
Specific Performance	**0.969**	**0.920**	0.827	0.855	0.168	0.193
(s.d.)	*(0.026)*	*(0.058)*	*(0.060)*	*(0.081)*	*(0.070)*	*(0.087)*
Damages vs. Specific Performance (p-val.)	0.603	0.716	0.328	0.644	0.054	0.895

Notes: *Selfish* refers to matching-groups where at most one buyer is classified as inequality-averse, while *Altruistic* refers to matching-groups where two or three buyers are classified as inequality-averse. All p-values are calculated using a two-sided Mann–Whitney U test. Numbers in bold indicate that the difference between Selfish and Altruistic groups, within the treatment, is statistically significant at the 5% level, according to a two-sided Mann–Whitney U test.

5. Limitations and Future Research

Our experimental design is deliberately tightly controlled to give the theory its best chance to succeed. While it may seem straightforward to expect subjects to follow theoretical predictions, this is rarely guaranteed—many theories fail even in highly favorable settings. Our findings do not preclude the possibility that future work may uncover differences across remedies. Rather, we offer a benchmark that future studies can build on by introducing additional complexity and realism.

Future theoretical work can further explore the comparative advantages of alternative remedies under costly renegotiation. With behavioral confounds minimized, future designs can focus more directly on issues of efficiency and fairness in contractual enforcement. The likelihood and magnitude of exogenous shocks to costs and benefits of performance are likely to differ across the spectrum of contractual relationships. For instance, when transaction costs can prevent efficient renegotiation, the choice of remedies may indeed be critical to minimize the expected social loss from inefficient performance or inefficient breach of contractual obligations.

A natural extension of our design would introduce asymmetric information—an essential feature of many real-world negotiations. In our current setting, all payoff-relevant information is common knowledge: in the Damages condition, the buyer observes the seller's increase in production costs; in the Specific Performance condition, the seller is informed of the buyer's updated valuation. A more realistic environment would allow parties to observe only that costs or outside options have changed, without precise quantification. It would be interesting to examine whether, and to what extent, renegotiation success and outcomes vary under such uncertainty, and whether these effects interact with the legal remedy.

Some features of our design warrant caution in interpreting the results. While we exogenously varied the contractual remedy, the framing may have lacked sufficient salience to elicit strong behavioral differences. Remedies were described in abstract terms, and payoffs were carefully calibrated to equalize efficiency and surplus across conditions. Prior evidence suggests that even in abstract contexts, endowment effects can arise (Isoni et al., 2011); however, one might expect stronger entitlement effects in contractual contexts involving tangible goods. Indeed, both Isoni et al. (2011) and Fehr et al. (2015) suggest that physical ownership may amplify the endowment effect relative to purely notional allocations. Beyond the use of real goods, future research could also consider the nature of the exchange: failure to transfer an existing good may elicit different reactions than failure to produce a new good or deliver a service. Should these dimensions prove consequential, they would offer empirical support for the mixed-remedy approach adopted in many legal systems.

Another feature of the design that may have attenuated treatment effects is the use of exogenously imposed prices. In our experiment, the contract terms—including price—were set by the experimenter, precluding participants from forming a genuine agreement or making an explicit promise. While such a constraint should not matter for fully rational agents, it may affect outcomes when participants differ in experience or when they attach psychological significance to the act of promising. These considerations are not merely hypothetical: in real-world contracting, parties often perceive promises as morally binding, and this may influence renegotiation behavior. An extension allowing for endogenous price formation would not only heighten the perceived authenticity of the agreement, but would also allow prices to reflect parties' expectations about the implications of different remedies. At the same time, initial prices would anchor subsequent renegotiation, potentially complicating the analysis of the successful renegotiation and the division of surplus.

Finally, it is worth noting that the high rate of successful renegotiation observed across both conditions may have introduced a ceiling effect. While this design feature preserved statistical power for analysing surplus division, future research might consider introducing frictions that generate greater variation in renegotiation outcomes.

6. Concluding Remarks

Parties enter into contracts to pursue mutually beneficial objectives. Remedies for breach of contract are necessary to ensure that parties can make credible commitments and rely on the agreements they have made. However, when market shocks occur, alternative breach remedies may be conducive to different outcomes. Specific performance may compel the fulfillment of a contract that has become inefficient (inefficient performance), while damage remedies might permit the breach of a contract that remains efficient (inefficient breach). According to Coase's (1960) theorem, if an exogenous shock changes the parties' respective costs and benefits under the original agreement, parties would bargain and renegotiate the existing contract to reach an efficient agreement. In the absence of transaction costs, situations of inefficient performance or inefficient breach would not occur, regardless of the applicable legal remedy.

In our incentivized experiment, we tested Coase's irrelevance proposition to verify if, in the shadow of distinct legal remedies, parties rationally renegotiate contracts that are no longer jointly desirable. Behavioral factors including inequality aversion and the endowment effect made us question if the type of legal remedy—specific performance versus damages—had any impact on the contracting parties' willingness and ability to renegotiate inefficient agreements and on the outcomes of such renegotiations. Parties seem to have a firm understanding of contractual efficiency and optimally renegotiate inefficient arrangements in the wake of market shocks. In line with Harrison and McKee (1985), we also find that prosocial concerns seem to play at best a moderate role in shaping the outcome of the bargaining process, which in our experiment approximates the Nash bargaining solution. Our experimental investigation of the effects of contract remedies on renegotiation dynamics finds no evidence of differences between remedies in this benchmark setting and is consistent with the predictions of Coase's hypothesis.

Statement

We hereby confirm that all research has been conducted in compliance with the IRB protocols of the institution receiving the grant. The project has received IRB approval on May 29, 2024, from the Ethics Committee of the University of Bologna (prot. N. 0156345).

References

Adams, R. (2022). Liquidated damages clauses: How do they work? *Texas Real Estate Research Center*. Retrieved from https://www.recenter.tamu.edu

Aivazian, V. A., Callen, J. L., & McCracken, S. (2009). Experimental tests of core theory and the Coase theorem: Inefficiency and cycling. *Journal of Law and Economics, 52*(4), 745–759. https://doi.org/10.1086/596562

Anidjar, L. Y. (2017). Globalization of the law of specific performance in contracts: Israel as a case study. *Journal of Comparative Law, 12*(1), 22–48. Retrieved from https://ssrn.com/abstract=3016338

Anidjar, L. Y., Katz, O., & Zamir, E. (2020). Enforced performance in common law versus civil law systems: An empirical study of a legal transformation. *American Journal of Comparative Law, 68*(1), 1–54. https://doi.org/10.1093/ajcl/avaa006

Ayres, I., & Talley, E. (1995). Solomonic bargaining: Dividing a legal entitlement to facilitate Coasean trade. *Yale Law Journal, 104*(5), 1027–1117. https://doi.org/10.2307/797059

Barton, J. H. (1972). The economic basis of damages for breach of contract. *Journal of Legal Studies, 1*(2), 277–304. https://doi.org/10.1086/467486

Bigoni, M., Bortolotti, S., Parisi, F., & Porat, A. (2017). Unbundling efficient breach: An experiment. *Journal of Empirical Legal Studies, 14*(3), 527–547. https://doi.org/10.1111/jels.12154

Birmingham, R. L. (1969). Breach of contract, damage measures, and economic efficiency. *Rutgers Law Review, 24*, 273–292.

Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review, 91*(1), 166–193. https://doi.org/10.1257/aer.90.1.166

Burlig, F., Preonas, L., & Woerman, M. (2020). Panel data and experimental design. *Journal of Development Economics, 144*, Article 102548. https://doi.org/10.1016/j.jdeveco.2020.102458

Calabresi, G., & Melamed, A. D. (1972). Property rules, liability rules, and inalienability: One view of the cathedral. *Harvard Law Review, 85*(6), 1089–1128. https://doi.org/10.2307/1340059

Cartwright, J., & Whittaker, S. (Eds.). (2017). The Code Napoléon Rewritten: French Contract Law after the 2016 Reforms. Hart Publishing.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics, 117*(3), 817–869. https://doi.org/10.1162/003355302760193904

Coase, R. H. (1960). The problem of social cost. *Journal of Law and Economics, 3*, 1–44. https://doi.org/10.1086/466560

Coase, R. H. (2006). The conduct of economics: The example of Fisher Body and General Motors. *Journal of Economics & Management Strategy, 15*(2), 255–278. https://doi.org/10.1111/j.1530-9134.2006.00100.x

Cooter, R. (1982). The cost of Coase. *Journal of Legal Studies, 11*(1), 1–33. https://doi.org/10.1086/467690

Cooter, R. D., & Ulen, T. S. (1997). *Law and economics* (2nd ed.). Addison-Wesley.

Dawson, J. P. (1959). Specific performance in France and Germany. *Michigan Law Review, 57*, 495–538. https://doi.org/10.2307/1286386

Depoorter, B., & Tontrup, S. (2012). How law frames moral intuitions: The expressive effect of specific performance. *Arkansas Law Review, 54*, 673–717. Retrieved from https://www.arizonalawreview.org/pdf/54-3/54arizlrev673.pdf

Durfee, E. N. (1935). *Remedies for breach of contract*. Hein & Co.

Edlin, A. S. (1998). Breach remedies. In P. Newman (Ed.), *The New Palgrave Dictionary of Economics and the Law*. Macmillan. https://doi.org/10.1007/978-1-349-74173-1_38

Engel, C. (2011) Dictator games: a meta study. *Experimental Economics,* 14, 583–610. https://doi.org/10.1007/s10683-011-9283-7

Engel, C., & Freund, L. (2017). *Behaviorally efficient remedies: An experiment* (No. 2017/17). Preprints of the Max Planck Institute for Research on Collective Goods. https://doi.org/10.2139/ssrn.2988653

Farnsworth, E. A. (1970). Legal remedies for breach of contract. *Columbia Law Review, 70*(7), 1145–1216. https://doi.org/10.2307/1121184

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191. https://doi.org/10.3758/BF03193146

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fauvarque-Cosson, B. (2017). The French contract law reform: A source of inspiration? *European Review of Contract Law, 13*(3), 275–294. https://doi.org/10.1515/ercl-2017-0018

Fehr, D., Hakimov, R., & Kübler, D. (2015). The willingness to pay–willingness to accept gap: A failed replication of Plott and Zeiler. *European Economic Review*, *78*, 120-128. https://doi.org/10.1016/j.euroecorev.2015.05.006

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics, 114*(3), 817–868. https://doi.org/10.1162/003355399556151

Fischbacher, U. (2007). z-Tree: Zurich Toolbox for Ready-Made Economic Experiments. *Experimental Economics, 10*(2), 171–178. https://doi.org/10.1007/s10683-006-9159-4

Fried, C. (1981). *Contract as promise: A theory of contractual obligation*. Oxford University Press.

Friedmann, D. (1989). The efficient breach fallacy. *Journal of Legal Studies, 18*, 1–24. https://doi.org/10.1086/468138

Goetz, C. J., & Scott, R. E. (1977). Liquidated damages, penalties and the just compensation principle: Some notes on an enforcement model and a theory of efficient breach. *Columbia Law Review, 77*(4), 554–594. https://doi.org/10.2307/1121823

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association, 1*(1), 114–125. https://doi.org/10.1007/s40881-015-0004-4

Harrison, G. W., & McKee, M. (1985). Experimental evaluation of the Coase theorem. *Journal of Law and Economics, 28*(3), 653–670. https://doi.org/10.1086/467104

Hillman, R. A. (2019). *Principles of contract law* (4th ed.). West Academic Publishing.

Hoffman, E., & Spitzer, M. L. (1982). The Coase theorem: Some experimental tests. *Journal of Law and Economics, 25*(1), 73–98. https://doi.org/10.1086/467008

Hoffman, E., & Spitzer, M. L. (1986). Experimental tests of the Coase theorem with large bargaining groups. *Journal of Legal Studies, 15*(1), 149–171. https://doi.org/10.1086/467807

Isoni, A., Loomes, G., & Sugden, R. (2011). The willingness to pay–willingness to accept gap, the "endowment effect," subject misconceptions, and experimental procedures for eliciting valuations: Comment. *American Economic Review, 101*(2), 991–1011. https://doi.org/10.1257/aer.101.2.991

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy, 98*(6), 1325–1348. https://doi.org/10.1086/261737

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives, 5*(1), 193–206. https://doi.org/10.1257/jep.5.1.193

Kalai, E. (1977). Proportional solutions to bargaining situations: Interpersonal utility comparisons. *Econometrica, 45*(7), 1623–1630. https://doi.org/10.2307/1913954

Kalai, E., & Smorodinsky, M. (1975). Other solutions to Nash's bargaining problem. *Econometrica, 43*(3), 513–518. https://doi.org/10.2307/1914280

Kaplow, L., & Shavell, S. (1996). Property rules versus liability rules: An economic analysis. *Harvard Law Review, 109*(4), 713–790. https://doi.org/10.2307/1342135

Kronman, A. T. (1978). Specific performance. *The University of Chicago Law Review, 45*(2), 351–382. Retrieved from https://www.proquest.com/scholarly-journals/specific-performance/docview/1301257383/se-2

Lando, O., & Beale, H. (Eds.). (2000). *Principles of European contract law: Parts I and II*. Kluwer Law International.

Lewinsohn-Zamir, D. (2013). Can't Buy Me Love: Monetary Versus In-Kind Remedies. *University of Illinois Law Review*, (1), 151-194. Retrieved from https://illinoislawreview.org/wp-content/ilr-content/articles/2013/1/Lewinsohn-Zamir.pdf

Lexology. (2020). An analysis of liquidated damages and contract rejection in bankruptcy. Lexology. Retrieved August 8, 2025, from https://www.lexology.com/library/detail.aspx?g=a204ddca-01b1-47b2-9089-afb999d6e56c

Mather, P. (1999). Financial covenants and related contracting processes in the Australian private debt market: An experimental study. *Accounting and Business Research, 30*(1), 29–42. https://doi.org/10.1080/00014788.1999.9728922

Medema, S. G. (2020). The Coase theorem at sixty. *Journal of Economic Literature, 58*(4), 1045–1128. https://doi.org/10.1257/jel.20191060

Nash, J. F. (1950). The bargaining problem. *Econometrica, 18*(2), 155–162. https://doi.org/10.2307/1907266

Nöldeke, G., & Schmidt, K. M. (1995). Option contracts and renegotiation: A solution to the hold-up problem. *The RAND Journal of Economics, 26*(2), 163–179. https://doi.org/10.2307/2555911

Parisi, F., Porat, A., & Bix, B. H. (2024). Opportunistic breach of contract. *Canadian Journal of Law & Jurisprudence, 37*, 199–230. https://doi.org/10.1017/cjlj.2023.20

Polinsky, A. M. (1980). Resolving nuisance disputes: The simple economics of injunctive and damage remedies. *Stanford Law Review, 32*(6), 1075–1112. https://doi.org/10.2307/1228549

Posner, R. A. (1977). *Economic analysis of law* (2nd ed.). Little, Brown.

Posner, R. A. (2009). Let us never blame a contract breaker. *Michigan Law Review, 107*(8), 1349–1364.

Rachlinski, J. J., & Jourden, F. (1998). Remedies and the Psychology of Ownership. *Vand. L. Rev.*, *51*, 1541–1582.

Rogerson, W. P. (1992). Contractual solutions to the hold-up problem. *Review of Economic Studies, 59*(4), 777–793. https://doi.org/10.2307/2297997

Rowan, S. (2017). The new French law of contract. *International and Comparative Law Quarterly, 66*(4), 805–841. https://doi.org/10.1017/S0020589317000252

Schwartz, A. (1979). The case for specific performance. *Yale Law Journal, 89*, 271–307. https://doi.org/10.2307/795838

Shavell, S. (2006). Is breach of contract immoral? *Emory Law Journal, 56*, 439–460. https://doi.org/10.2139/ssrn.868592

Shavell, S. (2009). Why breach of contract may not be immoral given the incompleteness of contracts. *Michigan Law Review, 107*(8), 1569–1581. Retrieved from https://repository.law.umich.edu/cgi/viewcontent.cgi?article=1325&context=mlr

Shiffrin, S. (2009). Could breach of contract be immoral? *Michigan Law Review, 107*(8), 1551–1568. Retrieved from https://repository.law.umich.edu/cgi/viewcontent.cgi?article=1324&context=mlr

Sidhu, D. S. (2006). The immorality and inefficiency of an efficient breach transactions. *The Tennessee Journal of Business Law, 8*, 61–100.

Sloof, R., Leuven, E., Oosterbeek, H., & Sonnemans, J. (2003). An experimental comparison of reliance levels under alternative breach remedies. *RAND Journal of Economics*, *34*(2), 205-222. https://doi.org/10.2307/1593714

Sloof, R., Oosterbeek, H., Riedl, A., & Sonnemans, J. (2006). Breach remedies, reliance and renegotiation. *International Review of Law and Economics, 26*(3), 263–296. https://doi.org/10.1016/j.irle.2006.11.002

Treitel, G. H. (1988). *Remedies for breach of contract: A comparative account*. Clarendon Press.

Treitel, G. H. (2003). The Law of Contract (11th ed.). Sweet & Maxwell.

Warkol, C. S. (1998). Resolving the paradox between legal theory and legal fact: The judicial rejection of the theory of efficient breach. *Cardozo Law Review, 20*(1), 321–354.

Wilkinson-Ryan, T., & Baron, J. (2009). Moral judgment and moral heuristics in breach of contract. *Journal of Empirical Legal Studies, 6*(2), 405–423. https://doi.org/10.1111/j.1740-1461.2009.01148.x

Williamson, O. E. (1985). Assessing contract. *Journal of Law, Economics, and Organization, 1*, 177–208. https://doi.org/10.1093/oxfordjournals.jleo.a036886

Zweigert, K., & Kötz, H. (1998). *An introduction to comparative law*. Oxford University Press.

Appendix A: Additional Tables

Table A.1: Balance table.

Variable	N	Damages (1) Mean/(Var)	N	Specific Performance (2) Mean/(Var)	N	Pairwise t-test (1)-(2) P-value
Female	144	0.528	144	0.535	288	0.906
		(0.251)		(0.251)		
Male	144	0.410	144	0.382	288	0.631
		(0.244)		(0.238)		
Age (in years)	144	24.681	144	24.722	288	0.924
		(15.673)		(11.740)		
Born in Italy	144	0.931	144	0.917	288	0.659
		(0.065)		(0.077)		

Table A.2: Panel regressions on heterogenous treatment effects.

	Renegotiation entry **(1)**	**Renegotiation success** **(2)**	**Surplus to sellers** **(3)**
Damages (d)	-0.014	0.025	-0.049
	(0.025)	(0.027)	(0.032)
Altruistic buyer	-0.048*	0.032	0.020
	(0.026)	(0.028)	(0.034)
Damages x Altruistic buyer	0.007	-0.006	0.047
	(0.037)	(0.039)	(0.047)
Period	0.005***	0.001	-0.005***
	(0.001)	(0.001)	(0.001)
Intercept	0.899***	0.817***	0.236***
	(0.019)	(0.024)	(0.023)
R-squared	0.089	0.007	0.098
N. obs.	680	680	670
Post estimation test			
Altruistic buyer + Damages x Altruistic buyer=0	0.121	0.349	0.043

Note: The post estimation test reports the p-value for the test: Altruistic buyer + Damages x Altruistic buyer = 0.

Appendix B: Instructions

Appendix B.1 Instructions for Damages Treatment

Instructions

Welcome. The purpose of this study is to investigate how people make decisions. From now until the end of the study,any communication with other participants is not allowed. If you have a question, please raise your hand and one of uswill come to your desk to answer it.

In this experiment, you will be able to earn money depending on your choices and the choices of the other participants.Upon completion of the study, the amount you earned will be paid to you via PayPal. Payments are confidential; no other participant will be told the amount you earned. All earnings are expressed in tokens, which will be converted toEuros at the end of the study at the rate of 1 Euro = 30 tokens.

This study is composed of two parts. We will now read instructions for Part 1. Instructions for Part 2 will be distributedat the end of Part 1.

Part 1

Roles and task. At the beginning of this part of the study, the computer will randomly assign you a "role": half of youwill be Sellers, the other half will be Buyers. Your role will remain fixed throughout the study.

Each Seller will be matched with one Buyer, and the two parties will be involved in a transaction.

- The Seller produces a good, and the expected production cost for each good is equal to 40 tokens.
- The Buyer expects to receive a benefit of 60 tokens from the purchase of the good.

The transaction is articulated in five phases.

Phase 1: Contract. The good is sold by the Seller to the Buyer, at a price of 50 tokens.

The two figures above show the screen for the Buyer (upper panel) and the Seller (lower panel). On the left side of theBuyer's screen, you can see that the benefit to the Buyer is 60 and the price is 50 so the Buyer's earnings from the contract will be 10 tokens (benefit minus price).

Likewise, on the left side of the Seller's screen, you can see that the price is 50 and the production cost is 40, so the Seller's earnings from the contract will be 10 tokens (price minus production cost).

In this phase of the experiment, there are no decisions to make, and the contract is binding for both parties.

Buyer

Period 1		Your Cumulative Earnings:	100

You have signed this contract.

YOU			SELLER		
Benefit:	60	-	Price:	50	-
Price:	50	=	Production Cost:	40	=
You Earn:	10		Your Seller Earns:	10	

Seller

Period 1		Your Cumulative Earnings:	100

You have signed this contract.

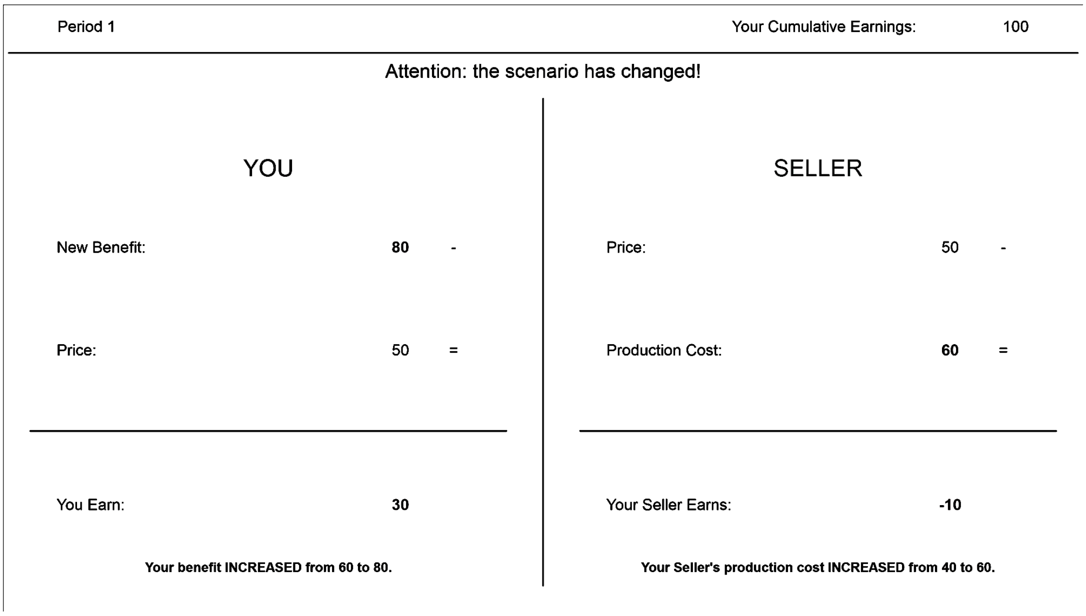
YOU			BUYER		
Price:	50	-	Benefit:	60	-
Production Cost:	40	=	Price:	50	=
You Earn:	10		Your Buyer Earns:	10	

Phase 2: The scenario may change. With probability 60% the conditions presented in phase 1 change.

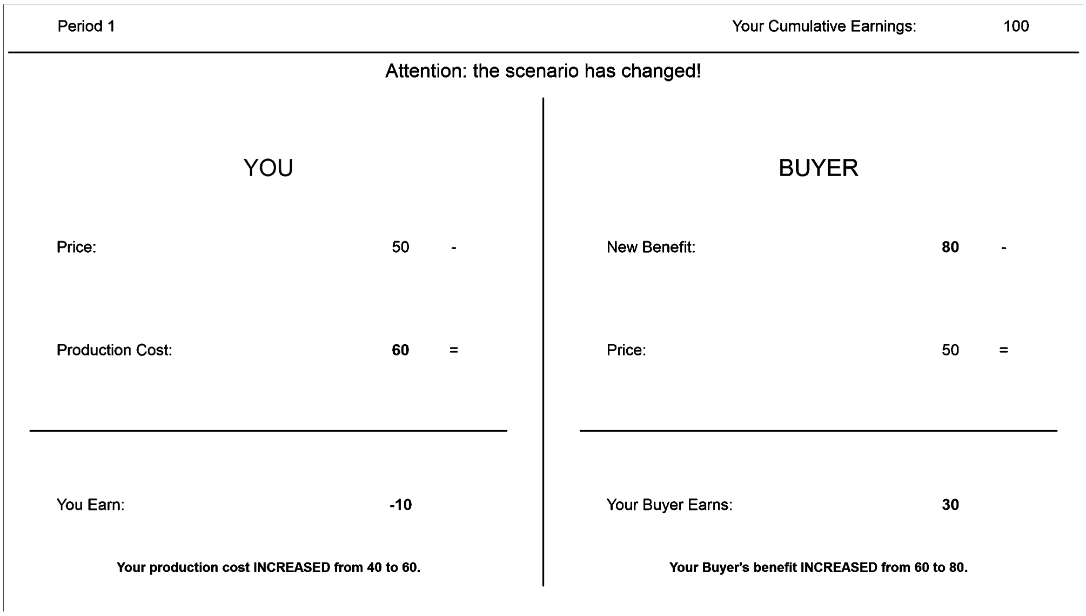
A. The production costs of the Sellers unexpectedly increase from 40 tokens to 60 tokens. In this case, if the con-tract is performed as promised, the Seller loses 10 tokens (price minus production costs).

B. The Buyer's benefit increases: that is, buying the good now yields a benefit of 80 tokens.

Buyer



Seller



Phase 3: Contract breach and renegotiation. If the scenario has changed, the Seller can unilaterally breach the contract, even without the consent of the Buyer.

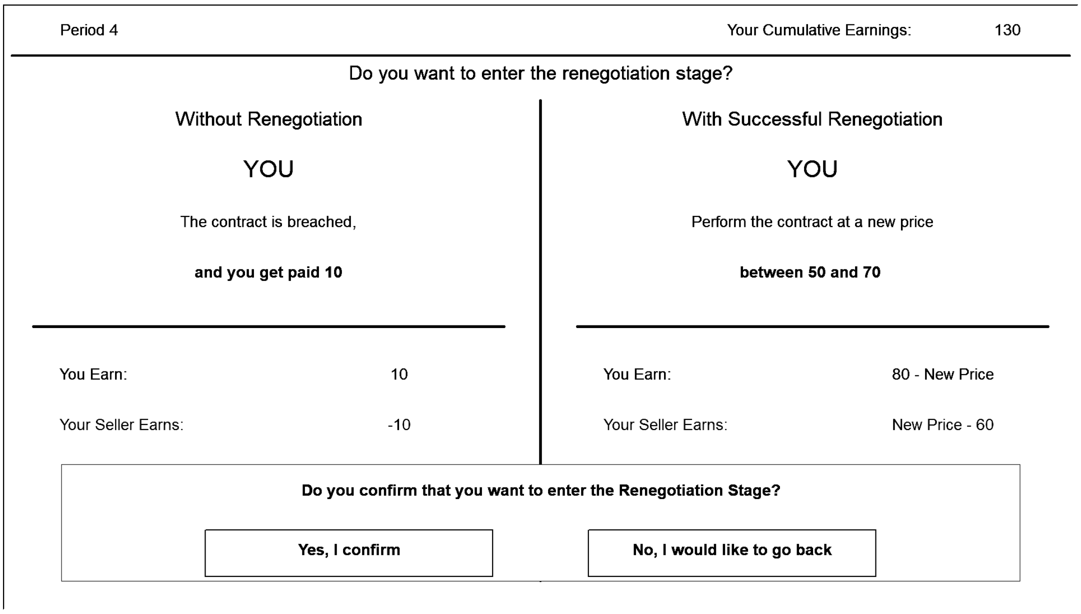
In this case the Seller will compensate the Buyer with 10 tokens (which is exactly the earnings a Buyer would have made by purchasing the good and keeping the good). Importantly, in this case the Buyer does not receive the good, and does not derive any benefit from it.

Alternatively, the Buyer and the Seller can renegotiate the contract, and agree on a new price.

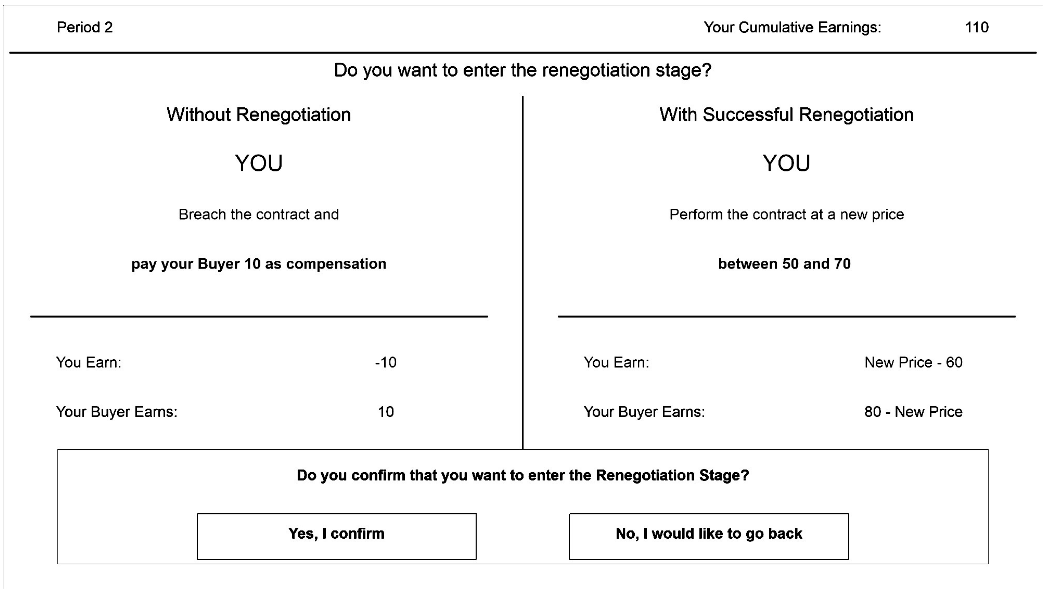
If the scenario has not changed, the Buyer and the Seller can perform the original contract and trade at a price of 50.Alternatively, they can renegotiate the contract, and agree on a new price.

The Buyer and the Seller simultaneously choose whether they want to renegotiate. Renegotiation is possible only if both choose to renegotiate.

Buyer

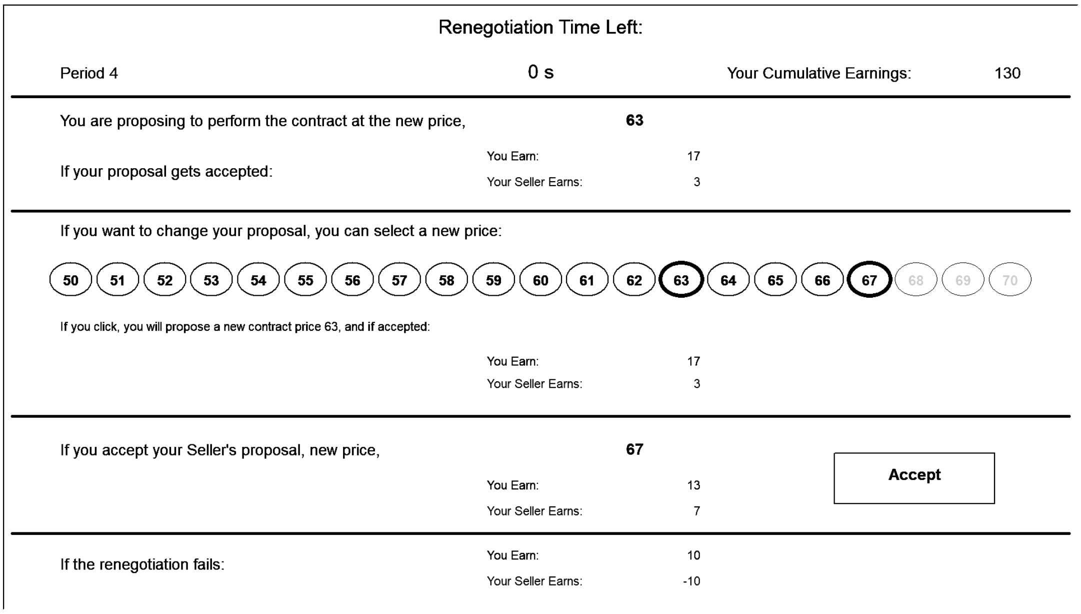


Seller

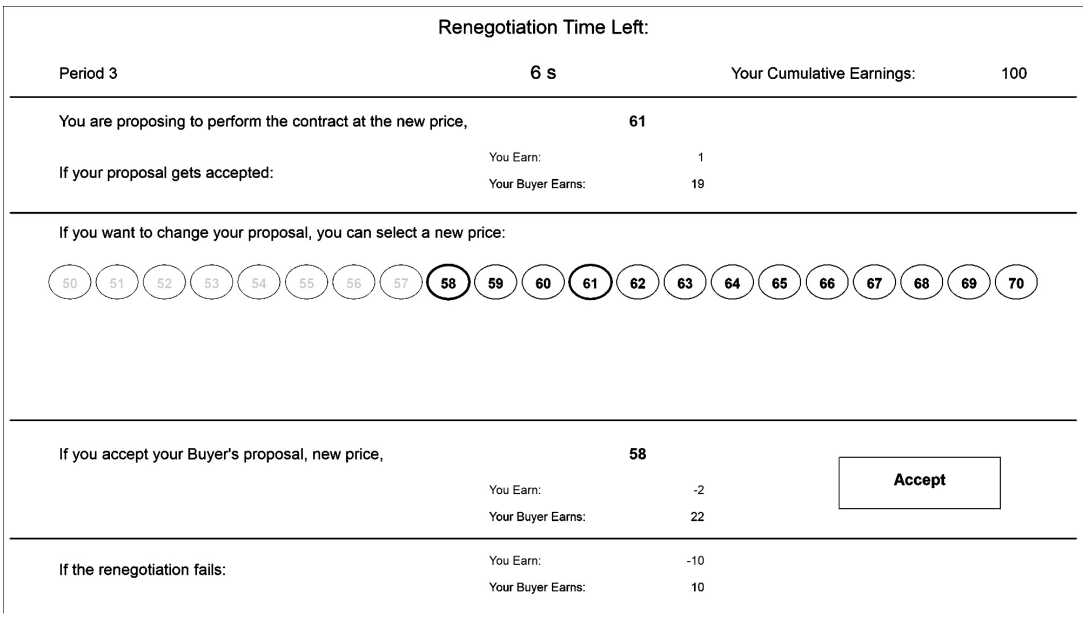


Phase 4: Renegotiation. If they choose to renegotiate, the Buyer and the Seller have 60 seconds to modify theoriginal contract.

Buyer



Seller



The Buyer and the Seller can propose a new price, initially included between 50 and 70. To propose a price, they can click on one of the numbers displayed in the middle of the screen. They can change their proposed price at any time, during the Renegotiation stage.

As soon as the Buyer proposes a price, this price is displayed in the lower part of the Seller's screen, who now has the possibility of accepting it, by clicking on the "Accept" button. Alternatively, the Seller can make a counteroffer by selecting a different price. The Seller will not be able to propose a price lower than the price proposed by the Buyer.

Similarly, as soon as the Seller proposes a price, this price is displayed in the lower part of the Buyer's screen, who now has the possibility of accepting it, by clicking on the "Accept" button. Alternatively, the Buyer can make a counterofferby selecting a different price. The Buyer will not be able to propose a price higher than the price proposed by the Seller.

If a price is accepted within the 60-second time limit, renegotiation succeeds and a new contract is performed at the new price.

If the Buyer and the Seller do not reach an agreement on the new price within the 60-second time limit, or they do notenter renegotiation, the outcome depends on whether there was a change in scenario.

- If there was no change in scenario, the original contract is performed.
- If there was a change in scenario, the original contract is breached, and the Seller will compensate the Buyerwith 10 tokens.

Phase 5: Outcome. Your screen displays a summary of what happened in Phases 2-4. The screen displays which variation in the initial conditions occurred in Phase 2, whether a renegotiation phase took place, and what was its outcome. On the lower part of the screen you can read your earnings, and the earnings of your counterpart.

Periods, groups, and private account. The task will be repeated for 25 periods. In each period the computer will form groups of two—one Seller and one Buyer—at random. You can see the number of the current period in the upper-left corner of the screen. In Phase 2 of each period, a shock might occur, with a 60% probability. The sequenceof the events is predetermined by the computer and cannot be influenced in any way by your previous actions.

At the beginning of the first period, an endowment of 150 tokens will be to your cumulative earnings. Per-period earnings will add up to your cumulative earnings too. In case you were to suffer a loss in a period, the tokens will be subtracted from your private account. Your cumulative earnings are always visible in the upper-right part of the screen.

To sum up:

- At the beginning of the experiment, you will be randomly assigned to the role of Seller or Buyer: the roles will remain fixed throughout the experiment.
- There will be 25 periods, and at the beginning of each period the computer will randomly match one Seller and one Buyer.
- In each period:
- The contract is signed;
- There can be a change in the initial conditions for the Seller and the Buyer: production costs increase for the Seller and the Buyer's benefit increases at the same time. The occurrence of this event is predetermined by the computer and does not depend in any way from your previous choices. You cannot know in advance the future sequence of events;
- The Seller can breach the original contract if the costs increase by paying damages;
- The Buyer and the Seller can renegotiate and agree on a new price;
- Renegotiation lasts 60 seconds.

The following table summarizes the outcomes in the 4 possible situations that can emerge in a period.

Change in scenario	Renegotiation	Contract	Seller's cost	Buyer's benefit	Price	Seller's earnings	Buyer's earnings
No	No	performed	40	60	50	10	10
No	Yes	performed	40	60	P' in 40-60	P'-40	60-P'
Yes	No	breached	60	80	—	-10	10
Yes	Yes	performed	60	80	P'' in 50-70	P''-60	80-P''

Earnings accumulate from period to period and are added to (or subtracted from) your private account.

We now ask you to answer a few questions, to verify that the instructions given so far are clear for everybody. The answers you give to these questions will not affect your earnings in any way.

Appendix B.2 Instructions for Specific Performance Treatment

Instructions

Welcome. The purpose of this study is to investigate how people make decisions. From now until the end of the study,any communication with other participants is not allowed. If you have a question, please raise your hand and one of uswill come to your desk to answer it.

In this experiment, you will be able to earn money depending on your choices and the choices of the other participants.Upon completion of the study, the amount you earned will be paid to you via PayPal. Payments are confidential; no other participant will be told the amount you earned. All earnings are expressed in tokens, which will be converted toEuros at the end of the study at the rate of 1 Euro = 30 tokens.

This study is composed of two parts. We will now read instructions for Part 1. Instructions for Part 2 will be distributedat the end of Part 1.

Part 1

Roles and task. At the beginning of this part of the study, the computer will randomly assign you a “role”: half of youwill be Sellers, the other half will be Buyers. Your role will remain fixed throughout the study.

Each Seller will be matched with one Buyer, and the two parties will be involved in a transaction.

- The Seller produces a good, and the expected production cost for each good is equal to 40 tokens.
- The Buyer expects to receive a benefit of 60 tokens from the purchase of the good.

The transaction is articulated in five phases.

Phase 1: Contract. The good is sold by the Seller to the Buyer, at a price of 50 tokens.

The two figures above show the screen for the Buyer (upper panel) and the Seller (lower panel). On the left side of theBuyer’s screen, you can see that the benefit to the Buyer is 60 and the price is 50, so the Buyer’s earnings from the contract will be 10 tokens (benefit minus price).

Likewise, on the left side of the Seller’s screen, you can see that the price is 50 and the production cost is 40, so the Seller’s earnings from the contract will be 10 tokens (price minus production cost).

In this phase of the experiment, there are no decisions to make, and the contract is binding for both parties.

Buyer

Period 4		Your Cumulative Earnings:	178

You have signed this contract.

YOU			SELLER		
Benefit:	60	-	Price:	50	-
Price:	50	=	Production Cost:	40	=
You Earn:	10		Your Seller Earns:	10	

Seller

Period 4		Your Cumulative Earnings:	142

You have signed this contract.

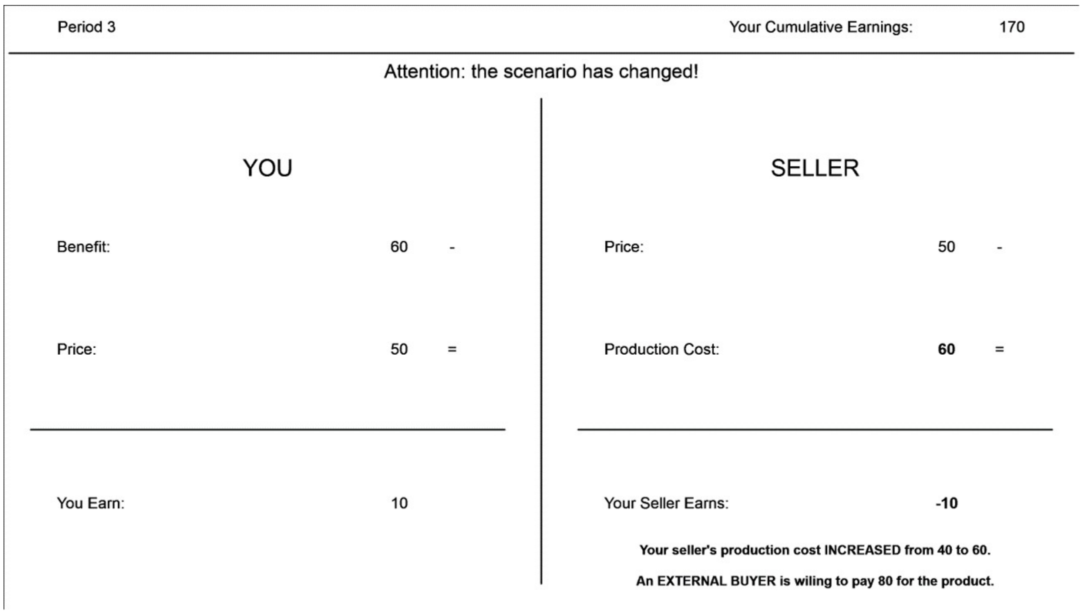
YOU			BUYER		
Price:	50	-	Benefit:	60	-
Production Cost:	40	=	Price:	50	=
You Earn:	10		Your Buyer Earns:	10	

Phase 2: The scenario may change. With probability 60% the conditions presented in phase 1 change.

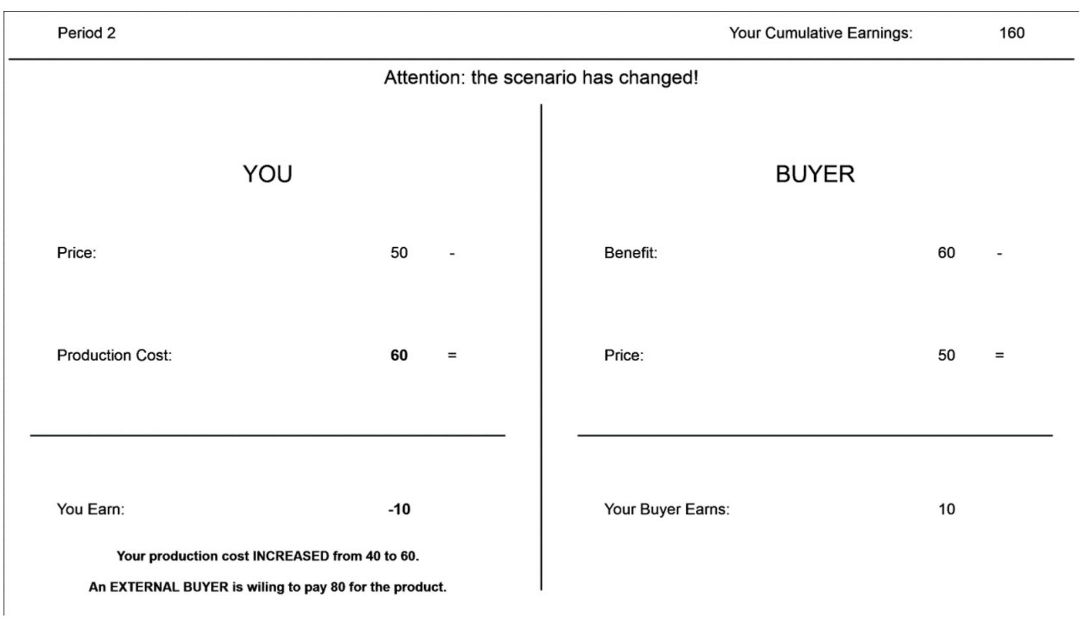
A. The production costs of the Sellers unexpectedly increase from 40 tokens to 60 tokens. In this case, if the con-tract is performed as promised, the Seller loses 10 tokens (price minus production costs).

B. The Seller gets an offer from another buyer; that is, if the current contract is canceled, the seller can sell theproduct to an external buyer for a price of 80.

Buyer



Seller

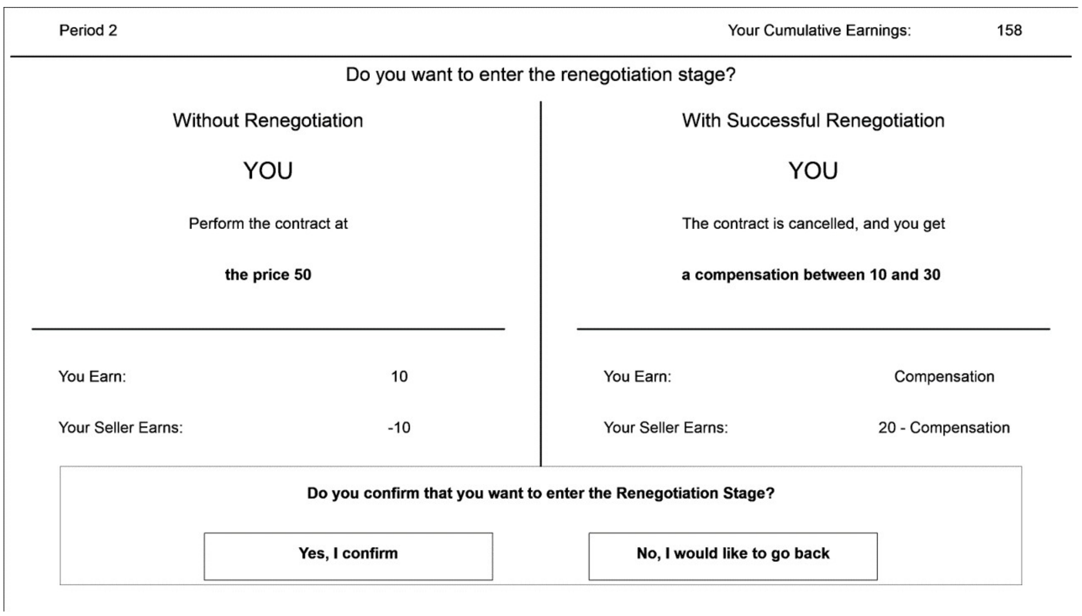


Phase 3: Contract breach and renegotiation. If the scenario has changed, the Seller cannot cancel the contract, without the consent of the Buyer.

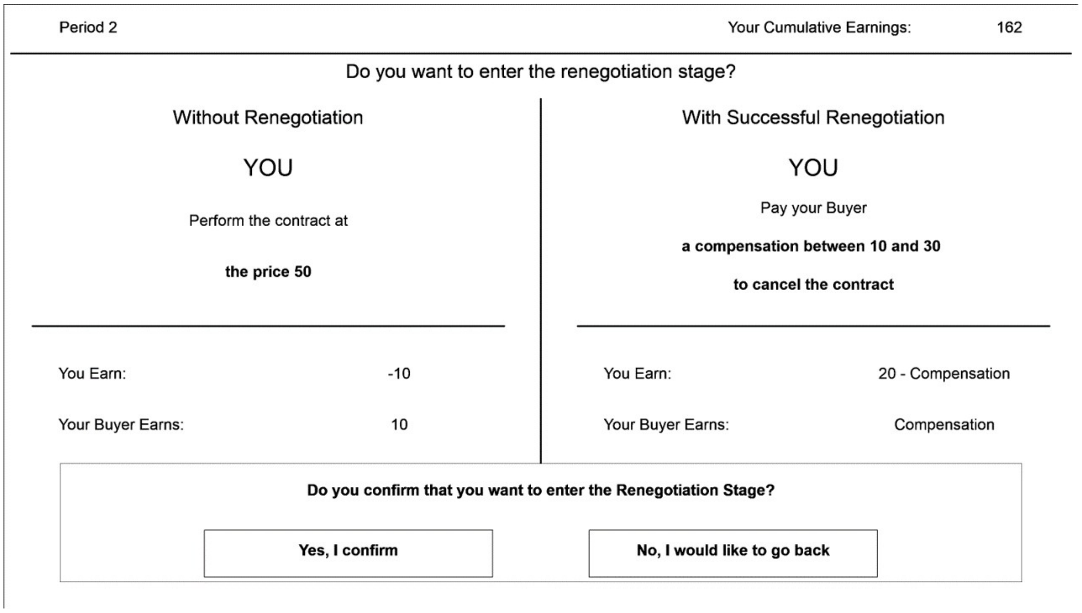
If the current is performed, the Seller will get a price of 50 and bear a production cost of 60, and the Buyer will pay a price of 50 and get a benefit of 60 from the product. Importantly, in this case the Seller cannot sell the product to the external buyer, and does not derive any benefit from this offer.

Alternatively, the Buyer and the Seller can renegotiate, and agree on a compensation from the Seller to the Buyer, to cancel the contract. In this case the Seller can sell the product to the external buyer, and get a price of 80.

Buyer



Seller

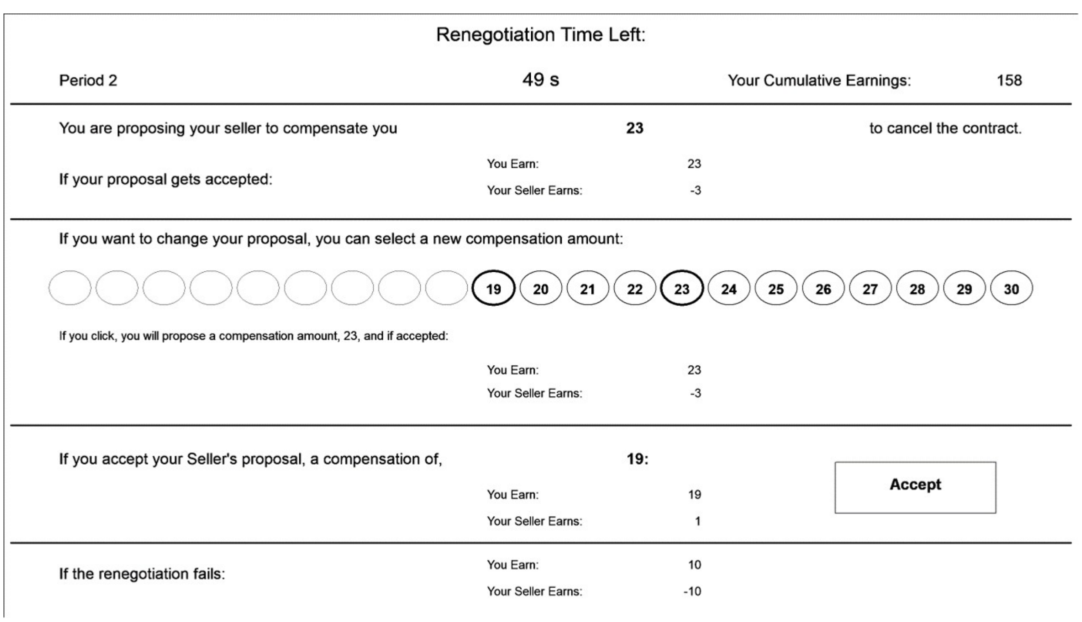


If the scenario has not changed, the Buyer and the Seller can perform the original contract and trade at a price of 50.Alternatively, they can renegotiate the contract, and agree on a new price.

The Buyer and the Seller simultaneously choose whether they want to renegotiate. Renegotiation is possible only ifboth choose to renegotiate.

Phase 4: Renegotiation. If they choose to renegotiate, the Buyer and the Seller have 60 seconds to agree on a compensation to cancel the original contract.

Buyer



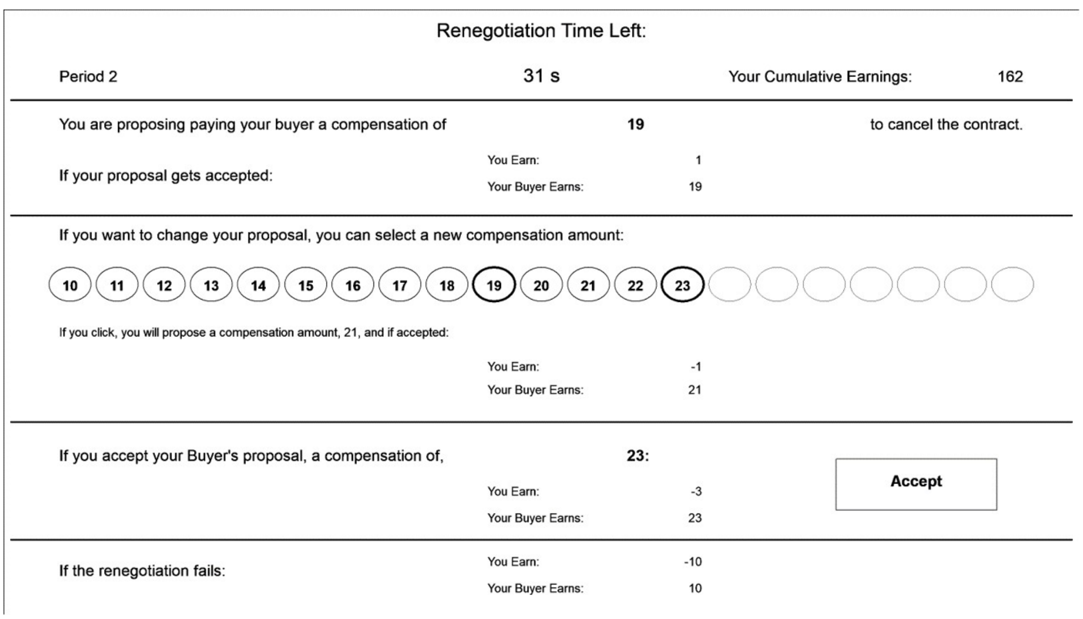
Seller

The Buyer and the Seller can propose a compensation, initially included between 10 and 30. To propose a compensation, they can click on one of the numbers displayed in the middle of the screen. They can change their proposed compensation at any time, during the Renegotiation stage.

As soon as the Buyer proposes a compensation, this is displayed in the lower part of the Seller's screen, who now has the possibility of accepting it, by clicking on the "Accept" button. Alternatively, the Seller can make a counterofferby selecting a different compensation. The Seller will not be able to propose a compensation higher than the one proposed by the Buyer.

Similarly, as soon as the Seller proposes a compensation, this is displayed in the lower part of the Buyer's screen, who now has the possibility of accepting it, by clicking on the "Accept" button. Alternatively, the Buyer can make a counteroffer by selecting a different compensation. The Buyer will not be able to propose a compensation lower thanthe one proposed by the Seller.

If a compensation is accepted within the 60-second time limit, renegotiation succeeds, the contract is canceled and theSeller transfers the agreed-upon compensation to the Buyer, and sells the good to the external buyer for a price of 80.



If the Buyer and the Seller do not reach an agreement on the compensation within the 60-second time limit, or they donot enter renegotiation, the original contract is performed and the outcome depends on whether there was a change inscenario.

- If there was no change in scenario, the Seller bears the original cost of 40.
- If there was a change in scenario, the Seller bears the increased cost of 60.

Phase 5: Outcome. Your screen displays a summary of what happened in Phases 2-4. The screen displays which variation in the initial conditions occurred in Phase 2, whether a renegotiation phase took place, and what was its outcome. On the lower part of the screen you can read your earnings, and the earnings of your counterpart.

Periods, groups, and private account. The task will be repeated for 25 periods. In each period the computer will form groups of two—one Seller and one Buyer—at random. You can see the number of the current period in the upper-left corner of the screen. In Phase 2 of each period, a shock might occur, with a 60% probability. The sequenceof the events is predetermined by the computer and cannot be influenced in any way by your previous actions.

At the beginning of the first period, an endowment of 150 tokens will be to your cumulative earnings. Per-period earnings will add up to your cumulative earnings too. In case you were to suffer a loss in a period, the tokens will be subtracted from your private account. Your cumulative earnings are always visible in the upper-right part of the screen.

To sum up.

1. At the beginning of the experiment, you will be randomly assigned to the role of Seller or Buyer: the roles willremain fixed throughout the experiment.
2. There will be 25 periods, and at the beginning of each period the computer will randomly match one Seller andone Buyer.
3. In each period:
    - The contract is signed;

      There can be a change in the initial conditions for the Seller: production costs increase and there is an external buyer willing to buy the product from the Seller at a higher price. The occurrence of this event is predetermined by the computer and does not depend in any way from your previous choices. You cannotknow in advance the future sequence of events;
    - The Seller cannot cancel the original contract, unless the Buyer agrees;
    - The Buyer and the Seller can renegotiate and agree on a compensation;
    - Renegotiation lasts 60 seconds.

The following table summarizes the outcomes in the 4 possible situations that can emerge in a period.

Change in scenario	Renegotiation	Contract	Seller's cost	Buyer's benefit	Price (P/P') or Compensation (c)	Seller's earnings	Buyer's earnings
No	No	performed	40	60	P=50	10	10
No	Yes	performed	40	60	P' in 40-60	P'-40	60-P'
Yes	No	performed	60	60	P=50	-10	10
Yes	Yes	canceled	60	60	c in 10-30	20-c	c

Earnings accumulate from period to period and are added to (or subtracted from) your private account.

We now ask you to answer a few questions, to verify that the instructions given so far are clear for everybody. The answers you give to these questions will not affect your earnings in any way.

Appendix B.3 Instructions for Dictator Game

Instructions

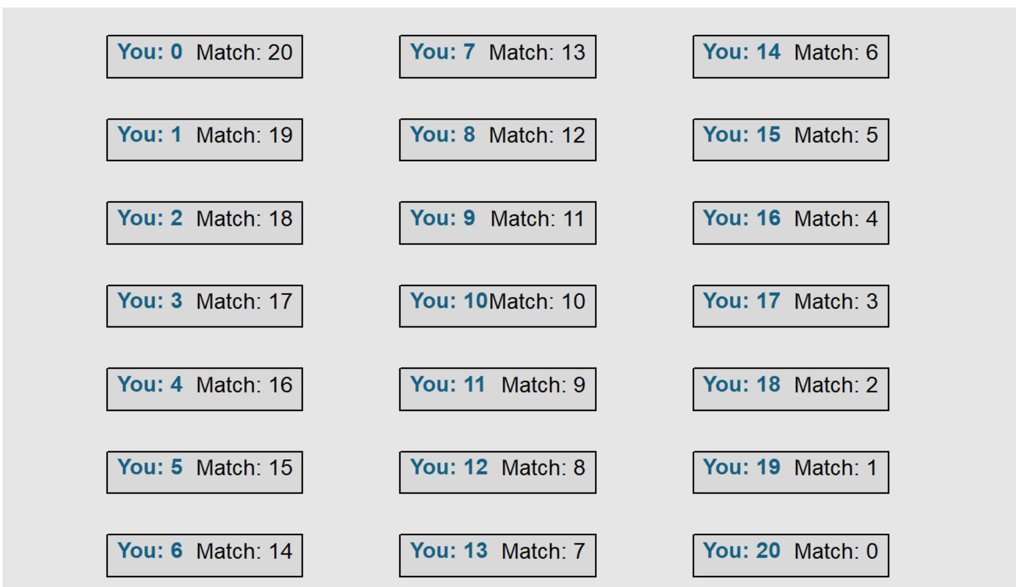
Part 2

Decision task. In this part you must decide how to allocate some tokens between you and another participant, whom we will refer to as ”your match”.

In the following Figure you can see the decision screen: each cell represents a possible allocation.

Please look at the first cell in the upper-left corner: in this distribution you get 0 tokens and your match gets 20. In the next allocation, you get 1 and your match gets 19. In the last allocation, in the bottom-rightcorner, you get 20 and your match gets 0.

To select the favorite allocation, you have to press on the desired cell and confirm your choice.



Groups and earnings. The computer will randomly form pairs. In each pair, the choice of one of the two participants will be randomly selected, and implemented. Hence, the implemented choice could be yours,or it could be the choice of the other person in your pair, that is, your match.

At the end of the study, you will know whether the selected choice is yours or your match’s, and you will know your earnings from Part 2. These will then be added to your cumulative earnings in Part 1 and converted into Euros to determine your final payment.

Appendix B.4 Questionnaires

Questionnaires

- **Were the instructions clear to you?**
    1) Not at all clear
    2) Not very clear
    3) Somewhat clear
    4) Very clear
- **Gender**
    1) Female
    2) Male
- **What is your age?**
- **Were you born in Italy**
    1) Yes
    2) No
- **Birthplace (if born in Italy)**
    1) Trentino AA, Veneto, Friuli VG
    2) Piemonte, Valle d'Aosta, Lombardia, Liguria
    3) Toscana, Umbria, Marche, Lazio
    4) Abruzzo, Molise, Puglia, Basilicata, Calabria, Campania
    5) Sicilia, Sardegna
- **Birthplace (if not born in Italy)**
    1) Europe
    2) Asia
    3) North America
    4) Center or South America
    5) Africa
    6) Oceania
- **In general, would you say that most people can be trusted or that you need to be very careful in dealing with people?**
    1) Most people can be trusted
    2) Need to be very careful
    3) Do not know

- **Do you think that most people would try to take advantage of you if they got a chance, or would they try to be fair? Please show your response on this card, where 1 means that "people would try to take advantage of you", and 10 means that "people would try to be fair":**

  1 [= People would try to take advantage of you], 2, 3, 4, 5, 6, 7, 8, 9, 10 [= People would try to be fair]

- **Please tell us, on a scale from 1 to 10 where 1 is never justified and 10 means always justified, whether you think it can always be justified, never be justified, or something in between: Claiming government benefits to which you are not entitled.**

  1 [=Never justified], 2, 3, 4, 5, 6, 7, 8, 9, 10 [=Always justified]

- **Please tell us, on a scale from 1 to 10 where 1 is never justified and 10 means always justified, whether you think it can always be justified, never be justified, or something in between: Cheating on taxes if you have a chance.**

  1 [=Never justified], 2, 3, 4, 5, 6, 7, 8, 9, 10 [=Always justified]

- **Are you generally ready to take on risks or you tend to avoid them? Please use this scale where 1 means "risk averse", while 10 means "ready to take risks".**

  1 [= Risk averse], 2, 3, 4, 5, 6, 7, 8, 9, 10 [= Ready to take risks]

- **Have you ever participated in experiments before?**

  1) No, never
  2) 1-2 times
  3) 3 times or more

- **Have you ever attended game theory courses?**

  1) Yes
  2) No

- **Have you ever attended law courses?**

  1) Yes
  2) No

Chapter 3

Break the Norm: An Experiment on Opinion Leaders and Homophily*

Xin Zhang†

Abstract: Prevailing social norms may persist even when most individuals privately prefer to violate them, due to uncertainty about others' attitudes. This paper studies whether an informed opinion leader—who knows the true distribution of preferences—can trigger norm change through public endorsement. Building on Gallice and Grillo (2025), I design a laboratory experiment in which both followers and a strategically motivated leader make decisions under uncertainty. I vary the leader's incentives and the social structure: in one condition, a neutral leader endorses violation only when individuals preferring to violate the norm are the majority; in another, an ideological leader benefits from successful norm-breaking even when such individuals are in the minority; and in a third, homophilous matching reduces perceived sanction risk. Results show that ideological leaders encourage more expression of norm-violating preferences in unfavorable environments, but behave similarly to neutral leaders when support is high. Homophily increases individuals' willingness to act on their preferences in some cases, though not consistently. These findings highlight how leadership incentives and social network structure jointly shape norm dynamics under uncertainty.

*The experiment has been initially pre-registered at AEA RCT on Dec. 17th, 2024. https://doi.org/10.1257/rct.10982-1.0. And the data collection was approved by the IRB of University of Bologna (IRB-0000990).

†PhD student in Economics, Department of Economics, University of Bologna

1 Introduction

Norm entrenchment can hinder the evolution of social norms, creating resistance to change despite legislation against them (Gulesci et al., 2023; Bénabou and Tirole, 2011; Bazzi et al., 2023). One of the sources of norm entrenchment is social sanctions, such as loss of reputation or social exclusion, which may persist even in the absence of formal legal punishment (Fehr and Fischbacher, 2004; Acemoglu and Jackson, 2017). Moreover, individuals, due to limited access to information, may hold incorrect beliefs about other people's attitudes toward certain social norms (Bénabou and Tirole, 2006; Ellingsen and Johannesson, 2008; Karlan and List, 2007). A possible scenario occurs when the majority of society actually prefers violating the prevailing social norm. However, due to fear of social sanctions, no one is willing to reveal their preference or act against the norm, leading to norm inertia. In this way, providing individuals with information about societal preferences regarding the norm may accelerate norm entrenchment (Bursztyn et al., 2020a).

Who can take on the role of transmitting the correct information to individuals? The person needs to have access to societal information and broad media coverage. One potential candidate is the opinion leader, which is the focus of this paper. This leads to the next question: Can opinion leaders help break the norm? And will they move in the socially optimal direction? A notable example is Desmond Tutu, who, in the 1990s, publicly supported women's rights in Africa, challenging the prevailing norm. This action, from current judgement, is more beneficial than harmful to the society. However, it is essential to recognize that social welfare might not be the sole motivation driving an opinion leader's endorsement. Factors such as ideology and concerns about popularity loss may heavily influence their decisions. Consider the case of Donald Trump, who, in 2020, discouraged mask-wearing, endorsing a norm violation that could have had adverse societal consequences during the COVID-19 pandemic.

Thus, this paper seeks to answer the question: "Under what circumstances will opinion leaders support the socially optimal outcome?" To investigate this, an experiment is designed to examine different circumstances based on three key variables: whether or not there is an opinion leader, whether the opinion leader has an ideological preference for violating the prevailing norm, and whether there is homophily within the society.

This paper contributes to three strands of literature: norm persistence under pluralistic ignorance, the influence of opinion leaders in coordination environments, and signaling in games with second-order uncertainty. It introduces a laboratory coordination setting in which individuals privately prefer to violate a prevailing norm but refrain from doing so unless they believe that others share their preference and will act on it. An informed opinion leader can issue a strategic endorsement—a publicly observable message intended to influence behavior—with incentives that vary across treatments. The experimental design isolates the effects of the leader's ideological incentives and the group's interaction structure (random vs. homophilous) on norm change. More fundamentally, the study examines the dynamic logic by which the leader's endorsement shapes individuals' beliefs and actions, which in turn feed back into the leader's future strategy. This framework provides a clean empirical test of whether—and under what conditions—opinion leaders can help overcome norm entrenchment.

The experiment is built around a simple coordination environment with latent preferences. In each round, individuals choose whether to publicly reveal a preference that violates a prevailing norm, knowing that doing so is only beneficial if enough others also reveal. A central feature of the design is the presence of an informed opinion leader who observes the true distribution of preferences and can issue a one-time, costless endorsement. This endorsement is observed by all individuals and may influence their beliefs and actions. The leader faces a tradeoff: endorsing norm violation may help like-minded individuals coordinate, but doing so also risks reputational or ideological costs, depending on the treatment. Treatments vary the opinion leader's ideological incentives—whether they are neutral or prefer the norm to be violated—and the way individual payoffs are shaped by others' revealed preferences. Specifically, the experiment manipulates whether individuals are randomly matched with peers or more likely to interact with those holding similar preferences (homophily), which affects the perceived risk of revealing a norm-violating stance. This setup allows for a clean test of how strategic signaling and social structure interact to influence norm-breaking behavior under second-order uncertainty.

More broadly, the framework captures situations in which breaking a prevailing norm is individually costly but may generate benefits for others who are not directly involved in the action. In such cases, norm-breakers can play a pivotal role in initiating change even when they constitute only a minority. Historical examples include men in male-dominated societies advocating to extend voting rights to women, free citizens supporting the abolition of slavery, or political dissidents in authoritarian regimes demanding democratic reforms. In all these cases, those who first challenged the norm incurred social, legal, or material costs, yet their actions helped create conditions that ultimately benefited others who remained passive. Analogously, in the experimental environment, potential norm violators face losses when acting alone in large groups, but their willingness to reveal may nonetheless initiate shifts that benefit third parties outside the experiment's immediate structure.

The results show that opinion leaders can indeed facilitate norm change, but their effectiveness depends on both their ideological incentives and the group's social structure. In supportive environments—where a majority prefers norm violation—all leaders tend to endorse, regardless of treatment, and most individuals reveal their preferences. However, in more challenging environments—where norm-violators are in the minority—only leaders with an ideological motive to promote norm violation are consistently willing to endorse. This increases the rate at which individuals reveal their true preferences, especially when the group is homophilous. Homophily alone has limited impact, but when combined with ideological endorsement, it strengthens individuals' willingness to act on latent preferences. These findings suggest that both leadership incentives and social context are critical in determining whether latent dissent can translate into collective norm change.

The remainder of the paper is organized as follows. Section 2 reviews related literature on norm entrenchment, opinion leadership, and coordination under uncertainty. Section 3 presents the experimental design, including the theoretical structure motivating the treatments and the resulting hypotheses. Section 4 reports the main results. Section 5 concludes.

2 Related Literature

This study builds on and contributes to three main strands of literature. The first focuses on the evolution and persistence of social norms, especially when individuals have incentives to conceal their true preferences. The second examines the role of opinion leaders in shaping behavior and belief coordination. The third explores how individuals respond to signals and coordinate under uncertainty, drawing on both global games—where players receive noisy information about an unknown state—and signaling theory, which studies how informed agents strategically transmit private information. This structure helps situate the experiment within broader debates about norm change under uncertainty. In particular, it clarifies how a strategically motivated opinion leader can influence collective behavior when societies are locked in possibly inefficient equilibria due to second-order uncertainty.

A central theoretical reference for this study is Gallice and Grillo (2025), who develop a model in which an informed opinion leader chooses whether to endorse norm violation under uncertainty, balancing ideological motives and reputational costs. While my experiment adapts the core logic of their framework, it introduces a more targeted, laboratory-friendly version suitable for testing how leadership incentives and social structure affect norm dynamics. The connection and differences are elaborated in the next section.

2.1 Social Norm Evolution and Entrenchment

Social norms play a central role in shaping individual behavior and coordinating expectations within societies. Yet even when many individuals privately prefer to deviate from a prevailing norm, such norms often persist. This puzzle of "norm entrenchment" has been studied both theoretically and experimentally, with key contributions highlighting the roles of pluralistic ignorance, belief-based coordination, and reputational concerns.

Theoretically, norms have been modeled as equilibria sustained by local conformity and stochastic stability (Young, 2014). In such settings, widespread private opposition may not translate into behavior change because individuals lack common knowledge of others' preferences. More recent models emphasize the importance of belief misalignment and second-order uncertainty—that is, uncertainty about others' beliefs—which can trap societies in inefficient outcomes unless a credible signal enables coordinated deviation.

Experimental studies provide empirical support for this view. Andreoni et al. (2020) show that visible tipping points can trigger rapid norm change once a minority acts. Smerdon et al. (2020) find that informational interventions often fail to dislodge harmful norms, while Szekely et al. (2021) show that environmental shocks can shift cooperative behavior. Casoria et al. (2021) document how policy changes during the COVID-19 pandemic led to immediate shifts in perceived norms and compliance. Relatedly, Dimant et al. (2024) show that individuals selectively seek norm-related information that aligns with their own behavioral preferences, reinforcing belief rigidity.

A particularly relevant subset of studies focuses on situations in which private preferences already support norm violation, but behavior remains frozen due to coordination risks. Bursztyn et al. (2020a), for example, find that Saudi men significantly underestimate peer support for women working outside the home; correcting this misperception increases both supportive attitudes and job applications. Similarly, Bursztyn et al. (2020b) show that Trump's election reduced perceived social sanctions around xenophobic speech, triggering behavioral shifts. Boudreau et al. (2025) demonstrate how union leaders in Myanmar enabled collective action by mobilizing individuals whose preferences already favored action but remained latent due to second-order uncertainty.

The current study contributes to this literature by experimentally examining both whether and how opinion leaders can facilitate norm change. While prior field studies show that leadership can coordinate norm-breaking behavior when preferences are misaligned with prevailing norms, these settings conflate leadership with other contextual factors. In a controlled environment, I replicate the core coordination challenge and test whether an informed endorsement is sufficient to unlock behavioral change. More importantly, I identify key mechanisms by varying the opinion leader's ideological incentives and the degree of homophily in the social structure. This allows for a cleaner test of how leader credibility and network composition jointly shape norm dynamics under second-order uncertainty.

2.2 Opinion Leaders and Informational Influence

Understanding how individuals update beliefs and change behavior in response to others' actions or advice is central to many domains of economics. Opinion leaders—individuals who hold influence due to status, credibility, or access to superior information—can play a critical role in shaping beliefs and coordinating behavior, especially when norms are contested or uncertain.

From a theoretical perspective, models of strategic communication highlight how senders with private information may manipulate signals to gain influence. For instance, Song (2025) develops a model in which competing opinion leaders strategically distort signals to attract followers, and in which followers weigh credibility when deciding whether to act on endorsements. These models suggest that leaders' incentives—whether ideological, reputational, or instrumental—critically affect their impact on belief coordination.

Empirical work confirms that influence depends not only on what leaders say, but on how they are perceived. In field experiments, Ambler et al. (2019) find that peer influence often outweighs formal leadership in agricultural adoption decisions, particularly when formal leaders are perceived as socially distant. In lab settings, Büchel et al. (2020) show that randomly assigned leaders can shape group decisions, but that their effectiveness depends heavily on perceived competence and trustworthiness. Even randomly selected or overconfident leaders can sway behavior, especially when others lack better information.

Other experiments underscore the broader challenges of social learning in strategic environments. Eyster and Weizsäcker (2018) find that individuals often mislearn due to naïve inference, such as neglecting correlation in observed actions. Conlon et al. (2025) show that individuals underweight accurate peer information in coordination games, suggesting that central, credible signals may be necessary to overcome inertia. These findings highlight the informational and social constraints on leader effectiveness.

This study contributes to this literature by experimentally varying the opinion leader's ideological incentives and their position within the social structure. While existing studies document that leaders can influence group behavior, especially under uncertainty, less is known about how strategic motives or perceived social distance shape their credibility. By embedding a leader's endorsement into an environment characterized by coordination risk and latent preferences, this experiment isolates the conditions under which informational influence succeeds or fails.

2.3 Strategic Signaling and Coordination in Global Games

A third strand of literature relevant to this study combines insights from global games and signaling theory to understand how individuals behave under uncertainty and strategic interdependence. Both literatures examine decision-making in the presence of incomplete information, but focus on different mechanisms: signaling games study how informed agents strategically transmit private information (Crawford and Sobel, 1982; Spence, 1973), while global games explore how coordination emerges when agents receive dispersed, noisy signals about an unknown fundamental (Carlsson and van Damme, 1993; Morris and Shin, 2003). The present study draws on both: individuals face a coordination problem under second-order uncertainty, as in global games, but the key signal comes from an informed opinion leader whose strategic incentives resemble those in signaling models. This hybrid structure allows for a richer investigation of how signal credibility, sender motivation, and social structure shape behavior when private preferences are latent.

In standard global game frameworks, individuals must decide whether to take a risky action—such as attacking a currency, joining a protest, or withdrawing from a bank—based on a private signal about a fundamental state. The payoff from acting depends not only on this signal, but also on how many others choose to act. Carlsson and van Damme (1993) show that introducing a small amount of uncertainty about fundamentals can lead to a unique equilibrium, resolving indeterminacy in classic coordination games. Morris and Shin (2003) extend this logic to macroeconomic settings, such as currency crises and financial stability. The current experiment mirrors this logic: individuals must decide whether to reveal a norm-violating preference based on uncertainty about others' actions and the underlying distribution of preferences. However, unlike standard global games in which signals are exogenously noisy, the public signal here is endogenously generated by a strategic opinion leader. This design captures both the informational structure of global games and the strategic transmission problem at the core of signaling theory.

Experimental work has validated many predictions of global games. Heinemann et al. (2004) show that even modest payoff uncertainty can guide equilibrium selection in coordination games. Heinemann et al. (2009) develop a method to directly measure strategic uncertainty, while Szkup and Trevino (2015) explore how players endogenously choose whether to acquire information before acting. These studies highlight how uncertainty about fundamentals interacts with higher-order beliefs to shape aggregate outcomes.

Related experiments explore how signal structure and credibility affect coordination. Duffy and Feltovich (2006) find that both cheap talk and costly signals can facilitate coordination, though their effectiveness depends on the environment. Shurchkov (2016) demonstrates that even uninformative public signals can shift behavior, particularly in repeated settings. Ioannou and Makris (2019) compare coordination across different information structures—common knowledge, global games, and Poisson games—and show that beliefs about the distribution of information can significantly alter outcomes. These findings point to the importance of signal framing and inference, even when no sender is explicitly modeled.

Additional research emphasizes cognitive and behavioral constraints in updating beliefs in coordination settings. Schmidt et al. (2003) find that players often gravitate toward risk-dominant strategies in one-shot games, while Shurchkov (2013) shows how learning evolves across repeated exposure to strategic uncertainty. These studies suggest that individuals may rely on heuristics or social cues rather than fully rational updating, especially when feedback is limited.

The current study contributes to this literature by integrating a strategic signal into a multi-agent coordination game under second-order uncertainty. The public signal—an endorsement by an informed opinion leader—plays the role of a shared but potentially biased cue, allowing followers to update beliefs not only about the underlying state but also about others' likely behavior. By varying the leader's ideological bias and the degree of homophily in the social network, the experiment isolates how sender incentives and audience composition jointly shape coordination. This structure bridges insights from global games and signaling theory, while offering new evidence on how credibility, motivation, and social context interact to influence norm change.

3 Experimental Design

The experiment is designed to study whether, and under what conditions, a strategically motivated opinion leader can facilitate norm change in a setting characterized by pluralistic ignorance and second-order uncertainty. In many real-world environments, individuals may privately disagree with a prevailing norm but refrain from acting on their preferences due to fear of social sanctions or doubts about others' willingness to deviate. A credible signal—especially from an informed or high-status individual—can shift beliefs and coordinate behavior, but its effectiveness depends on the sender's motives and the surrounding social context.

The experimental environment reproduces this coordination dilemma in a controlled setting. Participants must decide whether to reveal a preference that violates a prevailing norm, knowing that doing so is only beneficial if others do the same. An informed opinion leader observes the true distribution of preferences and can issue a one-time public endorsement of norm violation. The experiment varies the leader's ideological incentives and the group's interaction structure to test how endorsement strategies and peer exposure shape willingness to act on latent preferences. This setup enables a clean identification of the mechanisms through which leadership and social composition interact to either reinforce or break norm entrenchment.

3.1 The Game

Each group consists of five strategic players: one Gray player (the opinion leader) and four Purple players (potential norm violators). In addition, a number of Yellow players (computer-controlled norm abiders) are present in each round but are non-strategic and always choose the norm-abiding action, Y. The number of Yellow players can be either 2 or 72, forming either a Small group (2 Yellow players, 20% probability) or a Large group (72 Yellow players, 80% probability).

In each round, the game unfolds as follows: (i) Nature determines the composition of the group, which in the experiment is implemented through a binary signal indicating whether the group size is Small or Large; (ii) the Gray player observes this information and decides whether to issue a public endorsement encouraging Purple players to reveal their color (choose P); (iii) Purple players observe the Gray player's decision and simultaneously choose their actions, either to reveal their color (choose P) or not (choose Y); (iv) Purple and Yellow players are randomly matched, and payoffs are realized.

For the Gray player, after observing the group size, the action space consists of choosing whether to publicly endorse Purple color revelation or to remain silent. The Gray player earns a payoff only if an endorsement is issued. This payoff depends on the realized group size and the number of Purple players who choose P. Conditional on group composition, the payoff from endorsement is increasing in the number of Purple players choosing P, with the exact mapping varying across treatments. This structure generates a strategic tradeoff: issuing an endorsement creates the possibility of a positive payoff relative to remaining silent, which yields a payoff of zero, but also exposes the Gray player to potential losses if too few Purple players respond. The treatment-specific payoff structures are summarized in Table 2 and described in Section 3.2.

For a Purple player, after observing the Gray player's decision, the action space consists of choosing whether to reveal their color by selecting P or not reveal the color and appear norm-abiding by selecting Y. After all Purple players have made their choices, each Purple player is randomly matched with another player, either a Purple or a Yellow, according to the treatment-specific matching rule, and payoffs are realized. A Purple player who chooses P earns a payoff of 270 tokens if matched with a player who also chooses P, and incurs a payoff of -30 tokens if matched with a player who chooses Y. A Purple player who chooses Y earns a fixed payoff of 0, regardless of the matched player's choice. The payoff structure is summarized in Table 1.

		The PURPLE's choice	
		P	Y
The match's choice	P	270	0
	Y	-30	0

Table 1: The PURPLE's Payoff Structure

This game form adapts the signaling environment proposed by Gallice and Grillo (2025) into a laboratory setting. In their model, a continuum of agents decides whether to express a privately held, norm-violating opinion, while an influential sender (opinion leader) provides an informative public signal based on private observations. Expression is risky under pluralistic ignorance, and both the sender and the audience face second-order uncertainty. The current game preserves these core elements: private preferences, costly expression, an informed opinion leader, and a public signal that may shift collective behavior.

Several simplifications are introduced to ensure experimental tractability. First, the population is discretized into small groups with fixed roles: one opinion leader and four potential norm violators. Second, the leader's signal takes the form of a binary endorsement rather than a continuous message. Third, actions and roles are described using neutral labels: revealing a norm-violating preference is implemented as revealing one's color and choosing P, while norm-abiding behavior corresponds to not revealing their color and choosing Y. Finally, while each round involves sequential decisions—where the opinion leader observes private information and decides whether to endorse, followed by potential norm violators deciding whether to reveal—feedback is limited across rounds: outcomes and payoffs are shown after each round, but the underlying group composition remains hidden for PURPLE participants. These design choices allow for a clean empirical test of the model's core logic while preserving key features of belief formation and strategic signaling.

3.2 Treatments

The experiment features three between-subjects treatment conditions—Baseline, IDEO, and HOMO—that vary along two dimensions: the opinion leader's ideological incentives and the matching structure used to determine how PURPLE players are paired for payoffs. These variations are designed to isolate how leader motivation and social structure jointly affect the credibility and impact of norm-violating endorsements.

In the Baseline treatment, the GRAY player earns a payoff only if she endorses; otherwise, her payoff is zero. The payoff from endorsement depends on the group size and the number of followers. When the group is SMALL, the payoff increases with the number of PURPLEs who follow—starting from a positive value even if only one PURPLE chooses P. However, when the group is LARGE, the payoff is always negative, even if some PURPLEs follow. As a result, endorsement is optimal only in SMALL groups under Baseline conditions.

The IDEO treatment introduces ideological incentives by adding a bonus to the GRAY player's endorsement payoff in LARGE groups, while the payoff structure in SMALL groups remains identical to Baseline. That is, the GRAY earns a positive payoff in LARGE groups as long as at least one PURPLE follows the endorsement—though the benefit remains smaller than in the SMALL-group case. Consequently, endorsement is attractive in both group sizes in IDEO, provided at least one PURPLE chooses P. The leader's payoff continues to increase with the number of followers. Thus, IDEO differs from Baseline only in the GRAY's payoff structure in LARGE groups. The full payoff structure for the GRAY player in both treatments is summarized in Table 2.

Table 2: The Payoff Structure of the **Gray**s Choosing to Endorse

(1) Baseline

N. of PURPLEs choosing **P**	Group Size SMALL	Group Size LARGE
0	-4	-4
1	32	-3
2	36	-2
3	54	-1
4	60	0

(2) IDEO & HOMO

N. of PURPLEs choosing **P**	Group Size SMALL	Group Size LARGE
0	-4	-4
1	32	4
2	36	10
3	54	15
4	60	28

The HOMO treatment combines the same ideological incentives as in IDEO with a mod-

Table 3: The Expected Payoff of a **Purple** Player Choosing P

(1) Baseline & IDEO (2) HOMO

N. of other PURPLEs choosing **P**	Group Size SMALL	Group Size LARGE	N. of other PURPLEs choosing **P**	Group Size SMALL	Group Size LARGE
0	-30	-30	0	-30	-30
1	30	-26	1	32	-19.875
2	90	-22	2	94	-11.625
3	150	-18	3	156	-3.375

ified matching structure designed to introduce occasional homophily. In each round, there is a 5% probability that all PURPLE players will be matched exclusively with one another (internal matching), rather than being randomly matched with a mix of PURPLEs and Yellow players as in the other treatments. With the remaining 95% probability, the matching rule is identical to that used in Baseline and IDEO. This probabilistic matching variation increases the chance—albeit slightly—that a PURPLE player will be matched with a like-minded peer, thereby reducing the expected social cost of revealing a norm-violating preference. The corresponding expected payoffs for PURPLE players, depending on the number of other PURPLEs choosing P and the group size, are summarized in Table 3. Together with the opinion leader's ideological motive, this treatment creates a setting in which both the leader's incentives and the matching structure are more favorable to norm-violating behavior. All other features of the game—including timing, information structure, payoff rules, and belief elicitation—are held constant across treatments.

3.3 Procedures

At the beginning of experiment, participants are randomly assigned into a 5-participant group and to a role, either GRAY or PURPLE. The group assignment and roles remain fixed throughout the 25-round experiment. In each round, the group is joined by a random number of Yellow players (either 2 or 72) as stated in the Section 3.1.

The experiment comprises of two phases. Phase 1 consisted of five rounds in which no opinion leader was present and PURPLE players made decisions without receiving any endorsement. This phase was included to help participants familiarize themselves with the game structure. Data from these rounds are not used in the main analysis. The core game is played in Phase 2 for 20 rounds, where the Gray player can choose whether to issue a strategic endorsement encouraging norm violation. After observing the SMALL/LARGE group signal, the Gray player decides whether to send a public endorsement to the Purple players. This endorsement is a costless binary signal (endorse or remain silent) and is observed by all four Purples before they make their own decisions.

In addition to their strategic choices, participants also make incentivized guesses to capture their beliefs about others' actions and the state of the game. In each round, if the Gray player chooses to endorse, they are required to guess the number of Purple participants who will choose P in that round. They will receive a bonus of 20 tokens if their guess is exactly correct, and 0 otherwise.

Purple participants, after observing the Gray player's endorsement decision but before choosing their own action, are asked to report two beliefs. First, they provide a subjective probability (from 0% to 100%) that their group is LARGE. This belief is incentivized using a version of the matching probabilities scoring rule introduced by Danz et al. (2022), where participants earn 20 tokens if their reported probability aligns with a random draw in a way consistent with the actual state of the world. Second, each Purple guesses how many of the other three Purple participants will choose P. A correct guess earns 20 tokens; an incorrect one earns 0. These elicited beliefs capture participants' second-order uncertainty and expectations about coordination success, both of which are central to understanding how endorsement signals affect behavior.

To calculate final earnings, six rounds were randomly selected for each participant: one round from Phase 1[1] and five from Phase 2. Among the five Phase 2 rounds, three were used to determine payoffs based on participants' decisions, and two were used to evaluate their incentivized guesses[2]. The belief elicitation and decision payoffs were never evaluated

[1]The selected Phase 1 round contributed only a decision-based payoff. It was included to ensure that participants engaged seriously with all rounds, even though Phase 1 was not analyzed in the main results.

[2]For Gray players, the two selected rounds for belief elicitation both applied to the same type of guess:

in the same round, ensuring that choices and beliefs were not mechanically linked. This random selection procedure was clearly explained to participants in advance and ensured that all components of the game—including both strategic decisions and belief reporting—were made under incentive-compatible conditions. Final earnings were the sum of tokens earned across these selected rounds, converted to euros at a fixed exchange rate, 20 tokens = 1 euro.

The experiment was conducted in the BLESS laboratory at the University of Bologna using z-Tree (Fischbacher, 2007). A total of 140 student participants were recruited through ORSEE and randomly assigned to roles and treatments upon arrival. Upon entering the lab, participants scanned a QR code to access the information sheet and provided informed consent online. They were then seated and received printed instructions detailing the rules of the experiment, including game structure, incentives, and belief elicitation.

Participants completed a comprehension task before beginning the main task and were required to answer all questions correctly in order to proceed. Each session consisted of 25 rounds, during which participants remained in the same group. There were 28 groups in total: 4 in Baseline, 12 in IDEO, and 12 in HOMO. After the task, participants completed a short post-experimental questionnaire, which is the same in all treatments, including a set of background questions and a set of preference elicitation questions. The full questionnaire is available in the Appendix 5.2. Payments were made via PayPal. Participants received a €5 show-up fee plus performance-contingent earnings ranging from €5 to €56, with an average of approximately €15. The data were collected in December 2024. The pre-analysis plan has been submitted before the start of the data collection[3].

3.4 Theoretical Predictions

The theoretical predictions for all treatments are derived using backward induction. The equilibrium analysis primarily focuses on the strategy of Gray participants (the opinion

the number of Purple players choosing P (elicited only if the Gray endorsed). For Purple players, each of the two belief types—guessing the group size and guessing how many of the other three Purples would choose P—was incentivized once.

[3] Zhang, Xin. 2025. " Break the Norm: An Experiment on Opinion Leaders and Homophily." AEA RCT Registry. February 19. https://doi.org/10.1257/rct.15044-1.0. Initial registration date: December 17, 2024

leader), the choices of Purple participants (the potential violators), and the average payoff of Purple players, which serves as a measure of social welfare.

After observing the signal sent by the leader (Gray player), each follower (Purple) forms a posterior belief on the state of the world, that is on whether society is SMALL or LARGE. This corresponds to a beliefs about the probability of being paired with another Purple participant, denoted as $\beta(p|R)$. Here, p represents the actual probability of a Purple player being matched with another Purple player, and R denotes the Gray player's choice. There are two possible equilibria:

1. "Nobody chooses P": This is always an equilibrium since no Purple player has an incentive to deviate to choose P if all other Purple players choose Y.

2. "All Purples choose P": This equilibrium exists only when $\beta(p|R)$ is sufficiently high to ensure the expected payoff of choosing P is nonnegative. The condition for this equilibrium is:

$$\beta(p|R) \cdot 270 - (1 - \beta(p|R)) \cdot 30 \geq 0$$

$$\iff \beta(p|R) \geq 0.1.$$

If $\beta(p|R) < 0.1$, only the "Nobody chooses P" equilibrium exists. However, if $\beta(p|R) \geq 0.1$, both equilibria coexist, requiring Purple players to select one, so a coordination issue emerges. When $\beta(p|R) > 0.1$, the "All Purples choose P" equilibrium dominates in terms of payoff, as it yields a positive expected payoff. Additionally, when $\beta(p|R) > 0.2$, the "All Purples choose P" equilibrium risk dominates the "Nobody chooses P" equilibrium:

$$\beta(p|R) \cdot 50\% \cdot 270 - (1 - \beta(p|R) \cdot 50\%) \cdot 30 > 0$$

$$\iff \beta(p|R) > 0.2$$

This implies that for a Purple player, choosing P is more profitable, even if they assume that other Purple players choose P with only a 50% probability.

Previous literature indicates that when there is a conflict between the payoff-dominant and risk-dominant equilibria, the risk-dominant equilibrium is typically chosen (Schmidt et al., 2003; Février and Linnemer, 2006). Therefore, as shown in Figure 2, we assume that "All

PURPLEs choose P" will be selected as the equilibrium only when it is both payoff dominant and risk dominant (or risk indifferent). This condition holds when $\beta(p|R) \geq 0.2$.

Now consider the case in which there is no opinion leader (GRAY participants):

- In "Baseline" and "IDEO",

$$\beta(p) = 0.2 \cdot \frac{3}{5} + 0.8 \cdot \frac{3}{75} = 0.152 < 0.2$$

- In "HOMO",

$$\beta(p) = 0.2 \cdot (\frac{3}{5} * 0.95 + 0.05) + 0.8 \cdot (\frac{3}{75} * 0.95 + 0.05) = 0.1944 < 0.2$$

.

In both scenarios, $\beta(p) < 0.2$, and as a result, "Nobody chooses P" will be the equilibrium. Thus, without the participation of GRAY players (in Phase 1 across all treatments), the equilibrium is always "Nobody chooses P".

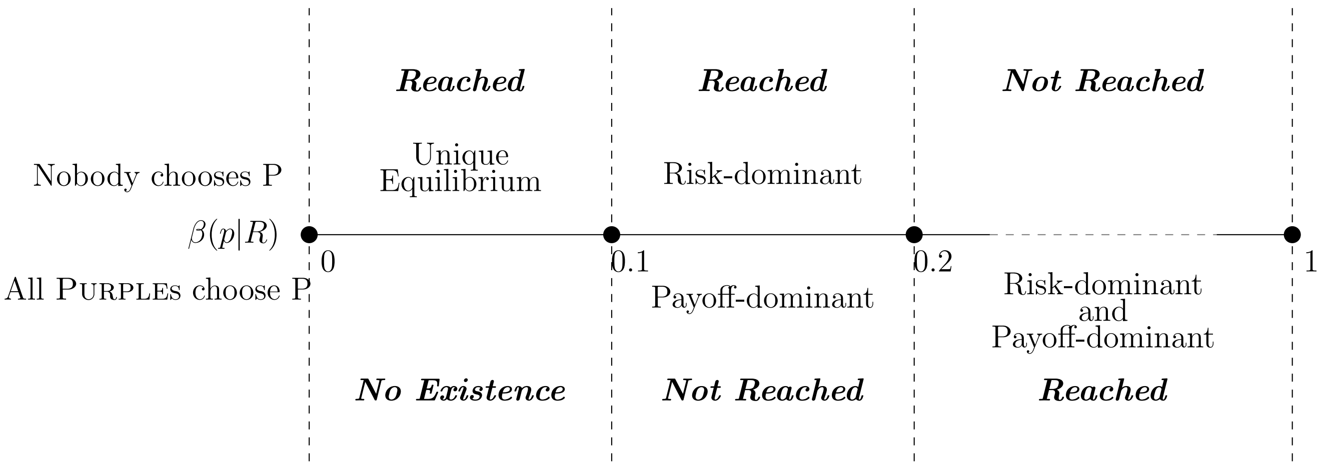


Figure 1: The Existence and Selection of the Equilibria for PURPLE Players

When the GRAY player (opinion leader) participates in the game, their task is to decide whether to encourage PURPLE players to choose P ($R = E$) or not ($R = NE$) based on the group size ("SMALL" - "S" or "LARGE" - "L"). Firstly, if they expect that "Nobody chooses P" will be reached by the PURPLE players with the endorsement, they won't choose $R = E$, since

$$\pi_{\text{GRAY}}|_{R=E \ \& \ n_P=0} = -4 < 0,$$

where n_p denotes the number of PURPLE players choosing P. Let $\eta^S \in [0,1]$ denote the probability that the GRAY participants choose $R = E$ when the group is SMALL, and $\eta^L \in [0,1]$ denote the probability that the GRAY participants choose $R = E$ when the group is LARGE. That is,

$$Pr(R = E|S) = \eta^S, \text{ and } Pr(R = E|L) = \eta^L.$$

Firstly, the GRAY participants need to make sure that the PURPLE players will select "All PURPLEs choose P" as the equilibrium when seeing the GRAY participant choosing E, that is

$$\beta(p|E) \geq 0.2 \Longleftrightarrow \begin{cases} \frac{5}{8}\eta^S \geq \eta^L, & \text{in "Baseline" and "IDEO"} \\ \frac{15}{16}\eta^S \geq \eta^L, & \text{in "HOMO"} \end{cases}$$

Proof: In "Baseline" and "IDEO",

$$\begin{aligned} \frac{Pr(R=E|S)Pr(S)}{Pr(R=E|S)Pr(S)+Pr(R=E|L)Pr(L)} * p_S + \frac{Pr(R=E|L)Pr(L)}{Pr(R=E|S)Pr(S)+Pr(R=E|L)Pr(L)} \cdot p_L &\geq 0.2 \\ \frac{\eta^S \cdot 0.2}{\eta^S \cdot 0.2 + \eta^L \cdot 0.8} * \frac{3}{5} + \frac{\eta^L \cdot 0.8}{\eta^S \cdot 0.2 + \eta^L \cdot 0.8} \cdot \frac{1}{25} &\geq 0.2 \\ \frac{0.6\eta^S + 0.16\eta^L}{\eta^S + 4\eta^L} &\geq 0.2 \\ \frac{5}{8}\eta^S &\geq \eta^L. \end{aligned}$$

In "HOMO",

$$\begin{aligned} \frac{Pr(R=E|S)Pr(S)}{Pr(R=E|S)Pr(S)+Pr(R=E|L)Pr(L)} * p_S + \frac{Pr(R=E|L)Pr(L)}{Pr(R=E|S)Pr(S)+Pr(R=E|L)Pr(L)} \cdot p_L &\geq 0.2 \\ \frac{\eta^S \cdot 0.2}{\eta^S \cdot 0.2 + \eta^L \cdot 0.8} * (\frac{3}{5} * 0.95 + 0.05) + \frac{\eta^L \cdot 0.8}{\eta^S \cdot 0.2 + \eta^L \cdot 0.8} \cdot (\frac{1}{25} * 0.95 + 0.05) &\geq 0.2 \\ \frac{0.62\eta^S + 0.352\eta^L}{\eta^S + 4\eta^L} &\geq 0.2 \\ \frac{15}{16}\eta^S &\geq \eta^L. \end{aligned}$$

Q.E.D.

Under the above constraint, if the group is SMALL, the payoff of the GRAY participants choosing E is

$$\pi_{\text{GRAY}}|_{S\&R=E \ \& \ n_p=4} = 60.$$

If the group is LARGE, the payoff of the GRAY participants choosing E is

$$\pi_{\text{GRAY}}|_{L\&R=E\ \&\ n_p=4} = \begin{cases} 0, & \text{in "Baseline"} \\ 28, & \text{in "IDEO" and "HOMO"} \end{cases}.$$

The GRAY participants' payoff maximization problem is

$$\max_{(\eta^S,\eta^L)} \eta^S \cdot \left[\pi_{\text{GRAY}}|_{S\&\ R=E\&\ n_p=4}\right] + \eta^L \cdot \left[\pi_{\text{GRAY}}|_{L\&\ R=E\&\ n_p=4}\right] \quad s.t.\ \beta(p|E) \geq 0.2.$$

Therefore, the GRAY player's equilibrium strategy is

$$(\eta^{S*}, \eta^{L*}) = \begin{cases} \eta^{S*} = 1, \eta^{L*} = 0 & \text{in "Baseline"}, \\ \eta^{S*} = 1, \eta^{L*} = \frac{5}{8} & \text{in "IDEO"}, \\ \eta^{S*} = 1, \eta^{L*} = \frac{15}{16} & \text{in "HOMO"}. \end{cases}$$

In summary, as displayed in the Table 4 below, the theoretical predictions on participants' CHOICE are:

- Phase 1: All PURPLEs choose Y.
- Phase 2:
  - GRAY:
    - If the group is SMALL, choose to endorse.
    - If the group is LARGE:
      - Baseline: choose to do nothing
      - IDEO: choose to endorse with a 62.5% probability and choose to do nothing with a 37.5% probability
      - HOMO: choose to endorse with a 93.75% probability and choose to do nothing with a 6.25% probability
  - PURPLE: choose P if the GRAY endorses while choose Y if the GRAY chooses to do nothing

The equilibrium payoff of a PURPLE participant from the CHOICE is presented in Table 5 below. In summary, the results indicate that the ideology of the GRAY participant (opinion

Table 4: Theoretical Predictions on GRAY and PURPLES' CHOICE

Treatment	Group Size	Phase 1		Phase 2	
		GRAY	PURPLE	GRAY	PURPLE
Baseline	SMALL	–	choose Y	Endorse	choose P
	LARGE			Not Endorse	choose Y
IDEO	SMALL			Encourage	choose P
	LARGE			Encourage with a 62.5% probability	choose P
				Not Endorse with a 37.5% probability	choose Y
HOMO	SMALL			Endorse	choose P
	LARGE			Endorse with a 93.75% probability	choose P
				Not Endorse with a 6.25% probability	choose Y

leader) reduces overall social welfare. This reduction occurs because the GRAY participant chooses to encourage norm violations even when they are aware that the group is LARGE. Conversely, homophily increases the probability that the GRAY participant provides encouragement in a LARGE group. However, homophily also mitigates the PURPLE participants' losses by reducing the negative payoff associated with revealing their color in a LARGE group, thereby increasing the equilibrium payoff for PURPLE participants.

Table 5: Theoretical Predictions on PURPLE's payoff from the CHOICE

Treatment	Phase 1			Phase 2		
	SMALL	LARGE	Overall	SMALL	LARGE	Overall
Baseline	0			150	0	30
IDEO				150	-18	15.6
HOMO				156	-3.6	28.5

3.5 Hypotheses

Based on the equilibrium analysis in Section 3.4, this subsection outlines a set of testable hypotheses concerning the behavior of both the opinion leader (GRAY) and the potential norm violators (PURPLEs), as well as the resulting implications for social welfare. The theoretical model yields clear comparative predictions across treatments regarding (i) the likelihood that GRAY endorses norm violation depending on group size and incentive structure; (ii) how PURPLEs respond to the presence or absence of endorsement; and (iii) how these dynamics influence average welfare, measured by the payoff PURPLEs receive from their CHOICE. The following hypotheses are designed to evaluate these mechanisms using experimental data, with a focus on treatment contrasts that isolate the effect of ideological incentives and homophilous matching. All hypotheses refer specifically to behavior and outcomes in Phase 2 of the experiment.

Hypothesis 1 GRAY*'s Strategic Endorsement Behavior:*

- *When the group is SMALL, the* GRAY *participant always endorse in all the treatments.*
- *When the group is LARGE, the* GRAY*'s endorsement rate in three treatments follows the ranking:*

$$\textit{Baseline (0\%)} < \textit{IDEO (62.5\%)} < \textit{HOMO (93.75\%)}$$

This hypothesis captures how endorsement behavior by the opinion leader responds to treatment-specific incentives. In SMALL groups, the endorsement is strictly optimal in all treatments, as it yields a guaranteed positive payoff when followed. In LARGE groups, however, endorsement becomes costly unless supported by incentives: the Baseline treatment offers no ideological reward and thus predicts zero endorsement; IDEO introduces ideological motivation, leading to endorsement with moderate probability (62.5%); and HOMO further improves coordination prospects through homophilous matching, increasing the expected success of endorsement and leading to a predicted endorsement rate of 93.75%. The resulting ranking provides a strong comparative benchmark for evaluating whether leaders adapt their behavior to strategic context.

Hypothesis 2 PURPLE*'s Norm Violating Behavior: Regardless of treatments or group size, the* PURPLE *participants will*

- *choose Y when the* GRAY *participant chooses not to encourage;*
- *choose P when the* GRAY *participant chooses to encourage.*

This hypothesis reflects the theoretical prediction that norm-violating behavior by PURPLE participants is highly responsive to the GRAY's endorsement decision. In the absence of encouragement, PURPLEs expect low coordination and risk incurring losses, making the dominant strategy to conform (choose Y). Conversely, when encouraged by the GRAY, the perceived probability of matching with another violator rises above the threshold needed to make choosing P both payoff-dominant and risk-dominant. As a result, PURPLEs are expected to follow the endorsement and reveal their norm-violating preferences. This behavior is predicted to hold across all treatments and group sizes.

Hypothesis 3 *Social Welfare (Average Payoff of* PURPLE *Participants):*

- *The average payoff from CHOICE, used as a proxy for social welfare, follows the ranking:*

$$IDEO < HOMO < Baseline$$

This hypothesis reflects the model's prediction that ideological endorsement increases norm violation even in environments where coordination is unlikely, thus lowering PURPLE players' average payoffs due to costly mismatches. However, when homophily is introduced (HOMO), the probability of successful coordination increases, reducing the downside of encouragement and partially restoring welfare. In contrast, in Baseline, where no encouragement occurs in unfavorable conditions, welfare remains higher due to avoidance of losses, even if positive change is suppressed.

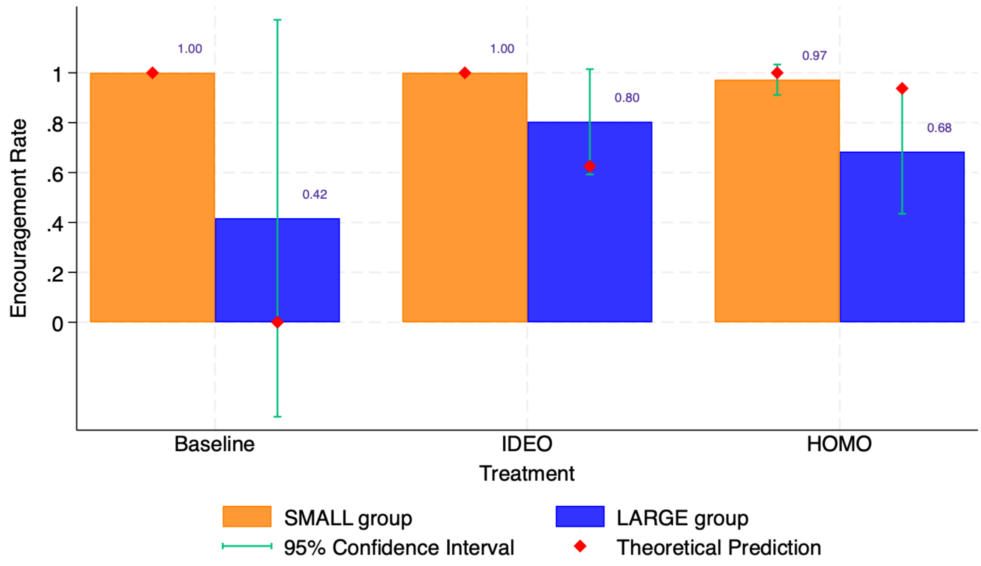
4 Results

This section presents the empirical results, testing the hypotheses derived from the theoretical model. The analysis proceeds in three parts. Section 4.1 examines the endorsement behavior of the Gray participants across treatments and group sizes, using both non-parametric comparisons and regression analysis. Section 4.2 turns to the norm-violating behavior of the Purple participants, conditional on the Gray's action, and assesses whether their responses align with the model's equilibrium predictions. Section 4.3 evaluates the implications for social welfare, measured by the average payoff of Purple participants from their revealing choices.

All analyses focus on the final 10 rounds of Phase 2, in order to reduce noise from early-round learning and adjustment. Unless otherwise noted, all hypothesis tests are conducted using the Mann–Whitney U test on group-level averages, with treatment group sizes of $n_{\text{Baseline}} = 4$, $n_{\text{IDEO}} = 12$, and $n_{\text{HOMO}} = 12$.

4.1 Gray's Endorsement Behavior

Figure 2 presents the average endorsement rate of the Gray participants in Phase 2, disaggregated by group size and treatment. Consistent with **Hypothesis 1**, endorsement in SMALL groups is nearly universal across all treatments (100%, 100%, and 97% separately), reflecting the high payoff associated with encouragement in favorable coordination environments. The p-value for a joint test of equality among the three rates is ($p > 0.1$, Kruskal–Wallis equality-of-populations rank test). In LARGE groups, endorsement is substantially lower in Baseline (mean = 0.42), considerably higher in IDEO (mean = 0.80), and somewhat lower in HOMO (mean = 0.68). The difference between Baseline and IDEO is marginally



Note: This figure displays the average endorsemnet rate of GRAY participants in the last 10 rounds under the 6 experimental conditions. The three sets of two bars (one orange and one blue) separatley diaplay the endorsement rate when the group is SMALL and LARGE in Baseline, IDEO, and HOMO treatments.

Figure 2: The GRAY participants' Encouragement Rate

significant [4] ($p = 0.08$, Mann–Whitney U test), whereas the difference between IDEO and HOMO is not statistically significant ($p = 0.40$).[5]

To formally test for treatment differences while controlling for individual preferences, we estimate the following linear probability model at the group-round level:

$$Endorse_{gt} = \alpha + \beta_1 \text{IDEO}_g + \beta_2 \text{HOMO}_g + \beta_3 \text{LARGE}_{gt} + \beta_4 (\text{IDEO}_g \times \text{LARGE}_{gt}) + \beta_5 (\text{HOMO}_g \times \text{LARGE}_{gt}) + \epsilon_{gt}$$

where the dependent variable is a binary indicator equal to 1 if the GRAY participant endorsed in group g, round t. Treatment indicators (IDEO, HOMO) capture average differences relative to BASELINE, and the interaction terms test whether treatment effects differ by group size.

[4] Throughout the paper, effects with p-values between 0.05 and 0.10 are described as marginally significant.

[5] Power calculations conducted prior to data collection indicated that the sample was sufficient to detect medium-sized effects. However, treatment contrasts in LARGE groups—such as between IDEO and HOMO—may remain underpowered.

Table 6: Regression Results on Endorsement Rate

	Endorsement rate	
	(1)	(2)
LARGE	-0.643***	-0.643***
	(0.141)	(0.151)
IDEO	-0.016	0.004
	(0.219)	(0.259)
HOMO	-0.081	0.079
	(0.219)	(0.264)
LARGE×IDEO	0.405***	0.331*
	(0.157)	(0.173)
LARGE×HOMO	0.351**	0.348**
	(0.157)	(0.168)
Preference control		YES
Constant	1.063***	1.006**
	(0.194)	(0.507)
R-squared	0.143	0.265
N. obs.	280	240

Note: 1. Standard errors in parentheses

2. *** p<0.01, ** p<0.05, * p<0.1

The regression results in Table 6 confirm the pattern observed in the raw endorsement rates. The interaction terms between treatment indicators and the LARGE group size dummy are both positive and statistically significant, indicating that endorsement behavior in LARGE groups is significantly more likely in both IDEO and HOMO compared to BASELINE. The coefficient for IDEO × LARGE suggests a substantial increase in the likelihood of endorsement when GRAY participants are motivated ideologically, even in less favorable coordination environments. However, the difference between the IDEO and HOMO is not statistically significant, implying that although HOMO increases endorsement rates relative to BASELINE, its incremental effect beyond IDEO cannot be distinguished with confidence given the current sample size.

4.2 Purple's Norm-Violating Behavior

Figure 3 displays the average revealing rate of Purple participants in Phase 2, conditional on whether the Gray participant chose to encourage or not. The results are broadly consistent with **Hypothesis 2**, although some deviations from theoretical predictions emerge.

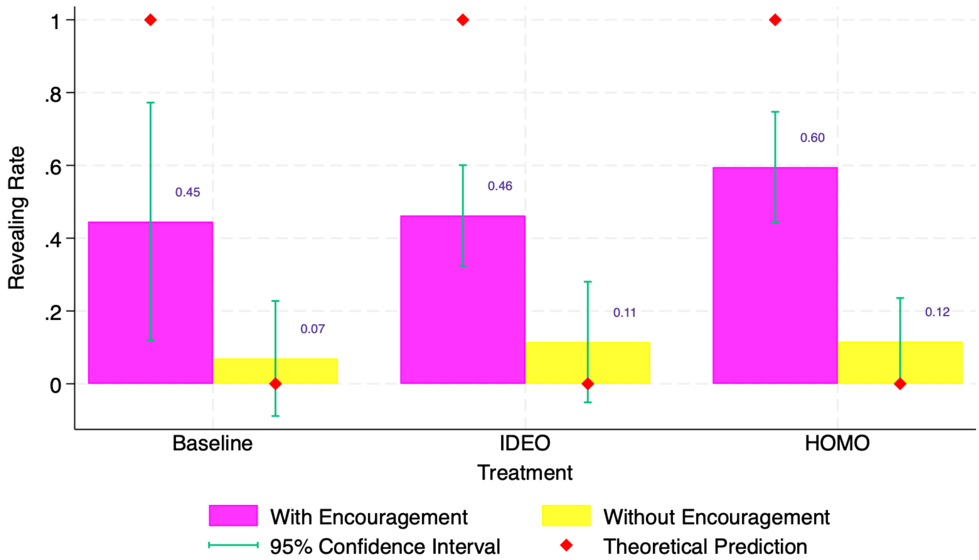
First, when the Gray chooses to encourage, the revealing rates are 45% in Baseline, 46% in IDEO, and 60% in HOMO—all significantly lower than the theoretically predicted 100% ($p < 0.01$). Pairwise comparisons show no significant difference between Baseline and IDEO ($p = 0.627$), but a significantly higher revealing rate in HOMO compared to IDEO ($p = 0.094$), suggesting that the homophilous matching in HOMO may have increased Purple participants' confidence in norm violation.

Second, when the Gray does not encourage, the revealing rates drop to 7%, 11%, and 12% in Baseline, IDEO, and HOMO, respectively. These rates do not differ significantly across treatments ($p > 0.50$), and are all consistent with the equilibrium prediction that Purple participants should refrain from revealing in the absence of encouragement.

To formally analyze how encouragement and treatment conditions shape norm-violating behavior, I estimate a linear regression model where the dependent variable is the *average revealing rate of* Purple *participants* in each group-round:

$$Reveal_{gt} = \alpha + \beta_1 \text{IDEO}_g + \beta_2 \text{HOMO}_g + \beta_3 Endorse_{gt} + \beta_4 (\text{IDEO}_g \times Endorse_{gt}) + \beta_5 (\text{HOMO}_g \times Endorse_{gt}) + \epsilon_{gt}$$

Regression results, reported in Table 7, confirm that endorsement by the Gray significantly increases the average revealing rate among Purple participants in the group (coefficient ≈ 0.34, $p < 0.01$). Furthermore, the interaction term HOMO $\times$ Encourage is positive and marginally significant, indicating that Purple participants in HOMO respond more strongly to encouragement compared to Baseline. In contrast, the interaction term for IDEO $\times$ En-



Note: This figure displays the average preference revealing (choosing P) rate of Purple participants in the last round under the six experimental conditions. The three sets of two bars (one pink and one yellow) separately display the revealing rate when the Gray participants in their group choose to encourage or not in Baseline, IDEO, and HOMO treatments.

Figure 3: The Purple Participants' Revealing Rate

courage is small and statistically insignificant. These results suggest that *homophilous matching*—rather than ideological framing alone—is more effective in eliciting norm-violating behavior through leader endorsement.

As a robustness check, Table 8 also reports the overall endorsement and revealing rates, averaged across all rounds in Phase 2 regardless of group size or encouragement. The general trend mirrors the conditional results: average endorsement is highest in IDEO (0.858) and HOMO (0.750) compared to Baseline (0.500), and the overall revealing rate is significantly higher in HOMO (0.463) than in Baseline (0.206; $p = 0.039$). These pooled statistics confirm that treatments fostering ideological motivation or homophilous matching increase norm violation at the aggregate level, even though individual behavior remains conditional on endorsemnt.

Table 7: Regression Results on Revealing Rate

	Revealing rate	
	(1)	(2)
Endorsed	0.336***	0.339***
	(0.085)	(0.085)
IDEO	-0.089	-0.088
	(0.111)	(0.112)
HOMO	0.045	0.049
	(0.104)	(0.106)
Endorsed×IDEO	0.160	0.157
	(0.106)	(0.107)
Endorsed×HOMO	0.169*	0.168*
	(0.100)	(0.100)
Preference Control	No	Yes
Intercept	0.038	0.112
	(0.086)	(0.215)
R-squared	0.362	0.362
N. obs.	280	280

Standard errors in parentheses.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

4.3 Efficiency and Social Welfare

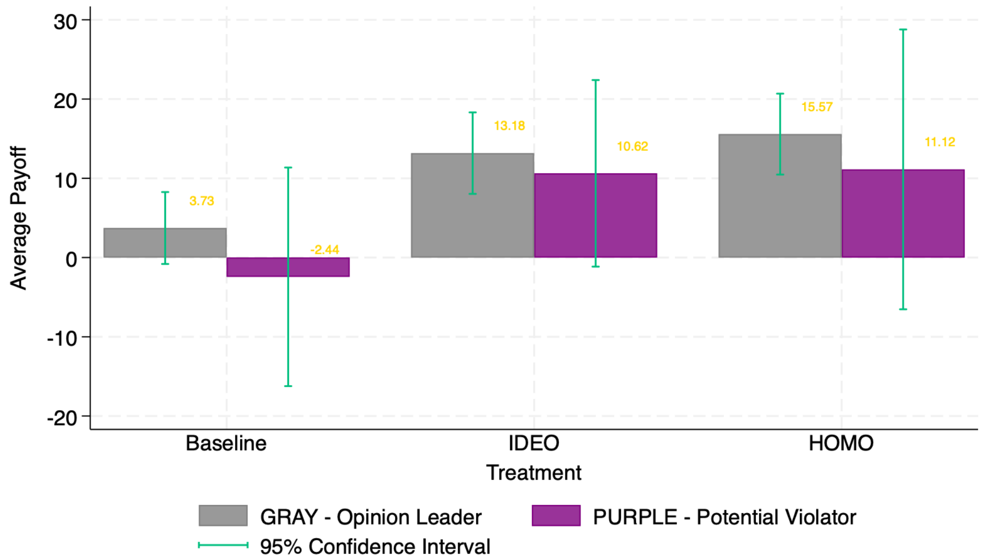
This section examines social welfare, which is proxied by the average payoff of Purple participants. These participants represent the potential norm violators, while other norm-abiders consistently choose the compliant action and therefore always receive a payoff of zero. As shown in Figure 4, there are no statistically significant differences in average Purple payoffs across the three treatments (p-value > 0.1). This may be partly due to limited statistical power, as the group-level design and sample size reduce the precision of treatment-level wel-

Table 8: Endorsement and Revealing Rates across Treatments

	Encouragement rate			Revealing rate		
	(L)	(S)	(All)	(E)	(NE)	(All)
Baseline	0.417	1.000	0.500	0.446	0.069	0.206
s.d.	(0.500)	(0.000)	(0.424)	(0.205)	(0.064)	(0.103)
IDEO	0.804	1.000	0.858	0.462	0.115	0.375
s.d.	(0.332)	(0.000)	(0.227)	(0.218)	(0.104)	(0.112)
HOMO	0.684	0.972	0.750	0.595	0.116	0.463
s.d.	(0.392)	(0.096)	(0.306)	(0.239)	(0.114)	(0.212)
Baseline vs. IDEO (p)	0.082*	.	0.082*	0.627	0.586	0.012**
Baseline vs. HOMO (p)	0.183	0.564	0.205	0.163	0.691	0.039**
IDEO vs. HOMO (p)	0.402	0.317	0.350	0.094*	0.913	0.213
N_{Baseline}	4	4	4	4	3	4
N_{IDEO}	12	12	12	12	4	12
N_{HOMO}	12	12	12	12	6	12

fare comparisons.

By contrast, the average payoff of Gray participants is significantly higher in the IDEO and HOMO treatments compared to Baseline (p-value < 0.05). This reflects the ideological payoff component that is unique to the ideological treatments and highlights a divergence between leader incentives and follower welfare.



Note: This figure displays the GRAY and PURPLE participants' average payoff in the last 10 rounds across three treatments. The three sets of two bars (one GRAY and one PURPLE) separately displays the average payoff of GRAY and PURPLE participants in Baseline, IDEO, and HOMO treatments.

Figure 4: Participants' Average Payoff

5 Conclusion

This study investigates how strategically motivated opinion leaders can influence norm change under second-order uncertainty. In environments where individuals privately prefer to violate a prevailing norm but hesitate due to potential sanctions and uncertainty about others' preferences, a leader's public endorsement can serve as a catalyst for change. The experiment varies both the leader's ideological incentives and the structure of peer interaction to identify how these factors shape collective outcomes.

More broadly, the findings speak to contexts in which taking the first step against a prevailing norm is personally costly but may yield broader social benefits. From historical struggles for civil rights to modern political dissent, successful norm change often begins with a minority willing to act despite sanction risks. This experiment illustrates how the presence of a strategically motivated leader—and a conducive social structure—can increase the likelihood of that initial coordination. By identifying the conditions that unlock such dynamics, the study offers insight into the micro-foundations of collective change.

The results highlight three key findings. First, when potential norm violators form a majority within their group, opinion leaders consistently choose to endorse norm violation across all treatments, and violators respond as predicted. Second, when norm violators are in the minority, endorsement becomes more selective and depends strongly on the leader's incentives. Ideologically motivated leaders are significantly more likely to encourage norm-breaking in these less favorable conditions, which increases norm-violating behavior among followers. While homophily marginally boosts the likelihood of norm violation, its combination with ideological motivation has the strongest effect—suggesting that endorsement is most effective when the audience is both similar and ideologically aligned. Third, although average payoffs among potential norm violators follow the predicted treatment ranking, these differences are not statistically significant. This may reflect limited statistical power and motivates further data collection.

To address this, I plan to conduct an additional wave of the experiment to improve the precision of treatment comparisons, particularly regarding welfare outcomes and belief-related dynamics. A promising direction for future research lies in unpacking the belief-updating process that underlies behavioral responses to endorsement. Although the current analysis focuses on observed choices, the experiment also elicited incentivized beliefs about others' actions and group composition. These data will be analyzed in future work to examine whether endorsement shifts expectations directly, or whether beliefs adjust only after behaviors begin to converge. Understanding these mechanisms will deepen our insight into how opinion leaders facilitate norm change under uncertainty.

A further extension concerns the role of moral content in norm adherence. The present design deliberately employs neutral language to isolate belief-based coordination and signaling mechanisms from moral framing effects. However, many real-world norms carry moral significance, and compliance may be driven not only by strategic considerations but also by intrinsic moral motivations. Testing the same endorsement mechanism in environments involving established norms with moral content—such as pro-social or civic behaviors—would allow an assessment of whether and how moral sentiments amplify, dampen, or qualitatively alter the impact of opinion leadership. Exploring this interaction represents a natural and important avenue for future research.

References

[1] Daron Acemoglu and Matthew O Jackson. Social norms and the enforcement of laws. *Journal of the European Economic Association*, 15(2):245–295, 2017.

[2] Kate Ambler, Susan Godlonton, and Maria P. Recalde. Follow the leader? a field experiment on social influence. *Journal of Economic Behavior & Organization*, 191:614–636, 2021.

[3] James Andreoni, Nikos Nikiforakis, and Simon Siegenthaler. Predicting social tipping and norm change in controlled experiments. NBER Working Paper No. 27310, 2020.

[4] Samuel Bazzi, Abel Brodeur, Martin Fiszbein, and Joanne Haddad. Frontier history and gender norms in the united states. Technical report, National Bureau of Economic Research, 2023.

[5] Roland Bénabou and Jean Tirole. Incentives and prosocial behavior. *American Economic Review*, 96(5):1652–1678, 2006.

[6] Roland Benabou and Jean Tirole. Laws and norms. Technical report, National Bureau of Economic Research, 2011.

[7] Laura Boudreau, Rocco Macchiavello, Virginia Minni, and Mari Tanaka. Leaders in social movements: Evidence from unions in myanmar. *American Economic Review*, 115(6):1975–2000, 2025.

[8] Konstantin Büchel, Ewoud Kloots, and Georg Weizsäcker. The strength of weak leaders: An experiment on social influence and social learning in teams. *Experimental Economics*, 23(1):249–278, 2020.

[9] Leonardo Bursztyn, Georgy Egorov, and Stefano Fiorin. From extreme to mainstream: The erosion of social norms. *American Economic Review*, 110(11):3522–3548, 2020.

[10] Leonardo Bursztyn, Alessandra L González, and David Yanagizawa-Drott. Misperceived social norms: Women working outside the home in saudi arabia. *American Economic Review*, 110(10):2997–3029, 2020.

[11] Hans Carlsson and Eric van Damme. Global games and equilibrium selection. *Econometrica*, 61(5):989–1018, 1993.

[12] Fortuna Casoria, Fabio Galeotti, and Marie-Claire Villeval. Perceived social norm and behavior quickly adjusted to legal changes during the covid-19 pandemic. *Journal of Economic Behavior & Organization*, 190:54–65, 2021.

[13] Christopher Conlon, Anandi Mani, Gautam Rao, Wendy Ridley, and Frank Schilbach. Not learning from others. Working paper, latest version January 2025, 2025.

[14] Vincent P. Crawford and Joel Sobel. Strategic information transmission. *Econometrica*, 50(6):1431–1451, 1982.

[15] Paolo Crosetto and Antonio Filippin. The "bomb" risk elicitation task. *Journal of Risk and Uncertainty*, 47:31–65, 2013.

[16] Eugen Dimant, Fabio Galeotti, and Marie-Claire Villeval. Motivated information acquisition and social norm formation. *European Economic Review*, 158:104563, 2024.

[17] John Duffy and Nick Feltovich. Words, deeds, and lies: Strategic behaviour in games with multiple signals. *The Review of Economic Studies*, 73(3):669–688, 2006.

[18] Tore Ellingsen and Magnus Johannesson. Pride and prejudice: The human side of incentive theory. *American Economic Review*, 98(3):990–1008, 2008.

[19] Erik Eyster and Georg Weizsäcker. An experiment on social mislearning. Working paper, 2018.

[20] Armin Falk, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. Global evidence on economic preferences. *The Quarterly Journal of Economics*, 133(4):1645–1692, 2018.

[21] Ernst Fehr and Urs Fischbacher. Social norms and human cooperation. *TRENDS in Cognitive Sciences*, 8(4), 2004.

[22] Philippe Février and Laurent Linnemer. Equilibrium selection: payoff or risk dominance?: the case of the "weakest link". *Journal of Economic Behavior & Organization*, 60(2):164–181, 2006.

[23] Andrea Gallice and Edoardo Grillo. Shifting social norms through endorsements. *Journal of Economic Behavior & Organization*, 218:29–48, 2025.

[24] Selim Gulesci, Sam Jindani, Eliana La Ferrara, David Smerdon, Munshi Sulaiman, and H Young. A stepping stone approach to norm transitions. 2023.

[25] Frank Heinemann, Rosemarie Nagel, and Peter Ockenfels. The theory of global games on test: Experimental analysis of coordination games with public and private information. *Econometrica*, 72(5):1583–1599, 2004.

[26] Frank Heinemann, Rosemarie Nagel, and Peter Ockenfels. Measuring strategic uncertainty in coordination games. *The Review of Economic Studies*, 76(1):181–221, 2009.

[27] Christos A. Ioannou and Miltiadis Makris. An experimental study of uncertainty in coordination games. *International Economic Review*, 60(2):751–799, 2019.

[28] Dean Karlan and John A List. Does price matter in charitable giving? evidence from a large-scale natural field experiment. *American Economic Review*, 97(5):1774–1793, 2007.

[29] Stephen Morris and Hyun Song Shin. Global games: Theory and applications. In Mathias Dewatripont, Lars Peter Hansen, and Stephen J. Turnovsky, editors, *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, volume 1, pages 56–114. Cambridge University Press, 2003.

[30] David Schmidt, Robert Shupp, James M Walker, and Elinor Ostrom. Playing safe in coordination games:: the roles of risk dominance, payoff dominance, and history of play. *Games and Economic Behavior*, 42(2):281–299, 2003.

[31] Simeon Schudy, Susanna Grundmann, and Lisa Spantig. Individual preferences for truth-telling. 2024.

[32] Olga Shurchkov. Coordination and learning in dynamic global games: Experimental evidence. *Experimental Economics*, 16(3):313–334, 2013.

[33] Olga Shurchkov. Public announcements and coordination in dynamic global games. *Journal of Behavioral and Experimental Economics*, 63:31–41, 2016.

[34] David Smerdon, Theo Offerman, and Uri Gneezy. 'everybody's doing it': on the persistence of bad social norms. *Experimental Economics*, 23(2):392–420, 2020.

[35] Yangbo Song. Social learning among opinion leaders. *Games and Economic Behavior*, 2025.

[36] Michael Spence. Job market signaling. *Quarterly Journal of Economics*, 87(3):355–374, 1973.

[37] Adriana Szekely, Axel Ockenfels, Gary Charness, and Arno Riedl. Evidence from a long-term experiment that collective risks change social norms and promote cooperation. *Nature Communications*, 12(1):1–11, 2021.

[38] Michal Szkup and Israel Trevino. Information acquisition in global games of regime change. *Journal of Economic Theory*, 160:387–428, 2015.

[39] H. Peyton Young. The evolution of social norms. *Annual Review of Economics*, 6(1):359–387, 2014.

Appendix

5.1 Instructions

5.1.1 Baseline treatment

Welcome to this study on economic decision-making.

The study consists of three phases:

in Phase 1 and phase 2 you will be matched with other participants and make some decisions.

in Phase 3 you will have to fill out a questionnaire.

At the end of the study, you will receive an amount of money which depends on your decisions and on the decisions of the other participants.

Instructions for Phase 1

Phase 1 lasts 5 rounds. In Phase 1, you will be part of a group which consists of **GRAY**, **PURPLE**, and **YELLOW** participants. **GRAY and PURPLE participants are real participants,** while **YELLOW participants are computerized**.

In each round of Phase 1, you will make a decision which may result in gaining tokens or losing tokens.

Who are the GRAY and PURPLE participants?

At the beginning of Phase 1, you will be randomly matched with 4 other real participants. One among the five of you will be randomly selected to take the role of the **GRAY** participant, while the 4 others will be **PURPLE** participants.

In all the 5 rounds of Phase 1,

- **your role will remain the same,** and
- you will be **grouped with the same other 4 real participants.**

Who are the YELLOW participants?

In your group there will also be some **computerized participants**. They will all take the role of **YELLOW** participants.

The **number** of **YELLOW** participants in your group is **randomly determined at the beginning of each round and can vary from one round to another**. This means that you will not know for sure the exact number of **YELLOW** (computerized) participants in your group. In each round, the number of **YELLOW** (computerized) participants can be

- **2** with **a 1-in-5 chance** (a 20% probability)
- **72** with **a 4-in-5 chance** (a 80% probability)

The size of your group

In each round, depending on the number of YELLOW (computerized) participants in your group, your group can be

- **"SMALL" - with 2** YELLOW participants
- **"LARGE" - with 72** YELLOW participants

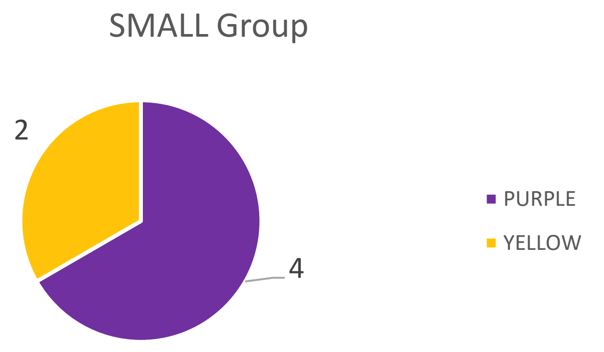


Figure 1: The composition of a "SMALL" group excluding a **GRAY** participant

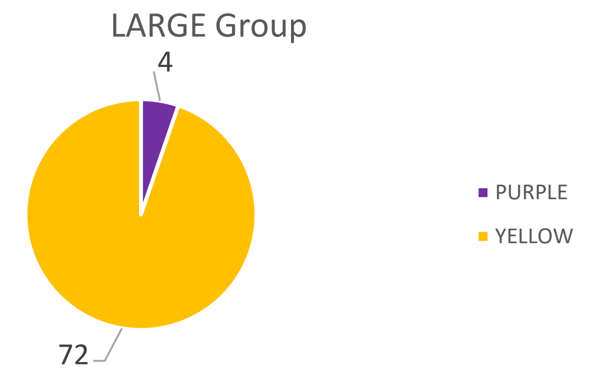


Figure 2: The composition of a "LARGE" group excluding a **GRAY** participant

Task of the PURPLE participants:

If you are a PURPLE participant, you will receive an endowment of 20 tokens at the beginning of each round.

In each round, your task is to decide whether to

- reveal your color and **choose P,**
- or do not reveal your color and **choose Y.**

After your decision, you will **be paired with another PURPLE or YELLOW (computerized)** participant in your group. This pairing is **randomly determined in each round and can vary from one round to another**. YELLOW **(computerized) participants always choose Y.**

Your payoff will depend on your choice and the choice of the participant you are paired with, as represented in the following table (in tokens).

		The other's choice	
		P	**Y**
Your choice	**P**	You **earn**: 270 The other **earns**: 270	You **lose**: 30 The other **earns**: 0
	Y	You **earn**: 0 The other **loses**: 30	You **earn**: 0 The other **earns**: 0

The table reads as follows:

1. If you choose to reveal your color, that is to **choose P**:
   - If you are paired with a participant **choosing P**, you will **earn 270 tokens**, and the other will earn 270 tokens;
   - If you are paired with a participant **choosing Y**, you will **lose 30 tokens**, and the other will earn 0 token.
2. If instead you **choose Y**, no matter who you are paired with, **you will earn 0 token.**
   - If the participant you are paired with chooses P, they will lose 30 tokens.
   - If the participant you are paired with chooses Y, they will earn 0 token.

Notice that **the chance that you are paired with a PURPLE participant** depends on the number of YELLOW (computerized) participants in your group.

If your group is **"SMALL"** (with **2** YELLOW participants),

- The chance that **you are paired with a PURPLE** participant is 3 in 5, that is **60%**
- The chance that **you are paired with a YELLOW** participant is 2 in 5, that is **40%.**

If your group is **"LARGE"** (with **72** YELLOW participants),

- The chance that **you are paired with a PURPLE** participant is 1 in 25, that is **4%**;
- The chance that **you are paired with a YELLOW** participant is 24 in 25, that is **96%.**

Task of GRAY participants:

If you are a GRAY participant, you will receive an endowment of 12 tokens at the beginning of each round.

In each round, you will be informed of your group size, either **"The Group is SMALL"** or **"The Group is LARGE"**.

After receiving the message, you need to guess

- **the number of PURPLE participants in your group that will choose P**

You will **earn**

- **20 tokens** if your guess is **correct**
- **0** if your guess is **incorrect**

Summary: sequence of actions

At the beginning of Phase 1, you will be randomly matched with 4 other real participants, and you will know your color, GRAY or **PURPLE**.

In each of the 5 rounds in Phase 1:

1. The number of **YELLOW** (computerized) participants in your group will be randomly determined: either 2 ("SMALL" group) with a 1-in-5 chance or 72 ("LARGE" group) with a 4-in-5 chance.
2. The **GRAY** participant in your group will receive a message about the size of your group. The **GRAY** participant will then guess how many **PURPLE** participants in your group will choose P.
3. Each **PURPLE** participant will decide to choose either **P** or **Y**. All **YELLOW** (computerized) participants will choose **Y**.
4. The **PURPLE** participants and **YELLOW** (computerized) participants in your group will be randomly paired.
5. All participants will know their payoff this round.

Earnings

At the end of experiment, the computer will **randomly select 1 of 5 rounds** in Phase 1 which will be relevant to determine your payment. Your earnings from Phase 1 will be equal to **your payoff in the randomly selected round**. Earnings will be converted into Euros at the rate of: **20 tokens = 1 euro**.

Instructions for Phase 2

Phase 2 lasts 20 rounds. In Phase 2, **you will be in the same group as in Phase 1, and your role will also be the same.** The situation you will face is also the same as in Phase 1, with a few differences, which will be described in detail in these instructions.

Tasks of the GRAY Participants

If you are a GRAY **participant,** you will receive an endowment of 12 tokens at the beginning of each round.

In each round, you will be informed of your group size, either **"The Group is SMALL"** or **"The Group is LARGE"**.

Choice

After receiving the message, you can choose:

- either to **publicly encourage the PURPLE participants to** reveal their color - **choose P**. In this way, all **PURPLE** participants in your group will see your encouragement,
- or to **do nothing**.

If you choose to **do nothing**, **your payoff will be 0**, no matter what other participants choose.

If instead you **choose to encourage PURPLE participants to choose P,** your payoff will depend on

- **your group size**,
- and on **the number of PURPLE participants choosing P in your group**.
  - **the higher** the number of **PURPLE** participants choosing P, **the higher** the payoff you will get.

The table below shows the exact payoff you would get if you choose to **encourage PURPLE participants to choose P.**

Total number of PURPLE participants choosing P	The GRAY participant's payoff if choosing to encourage PURPLE participants to choose **P**	
	"SMALL" group	"LARGE" group
0	**-4**	**-4**
1	**32**	**-3**
2	**36**	**-2**
3	**54**	**-1**
4	**60**	**0**

This implies that, for example, if you **choose to encourage PURPLE participants to choose P,** and in your group, there are **3 PURPLE** participants choosing **P** and **1 PURPLE** participant choosing **Y,**

a. If your group is **"SMALL"**, you will **earn 54** tokens;
b. If instead, your group is **"LARGE",** you will **lose 1** tokens.

Guess

And if you choose to encourage, you also need to make a guess of

- **the number of PURPLE participants in your group that will choose P**

You will **earn**

- **20 tokens** if your guess is **correct**
- **0** if your guess is **not correct**

Tasks of the **PURPLE** Participants

If you are a PURPLE participant, you will receive an endowment of 20 tokens at the beginning of each round.

Choice

In each round, your task is to decide whether to

- reveal your color and **choose P,**
- or do not reveal your color and **choose Y.**

The payoff structure is the same as in Phase 1.

Before making the decision, you will **first see the GRAY participant's choice,** whether they encourage all **PURPLE** participants to choose **P** or not, then you have to **make two guesses**.

Guess

Firstly, you **guess the chance that your group is LARGE in this round**. Your guess is a percentage probability from 0 to 100

- with 0 indicating 0-in-100 chance that your group is LARGE (i.e. you are sure that your group is SMALL),
- and 100 indicating a 100-in-100 chance that your group is LARGE (i.e. you are sure that your group is LARGE).

The number you provide is called ***Your Guess 1.***

You choose *Your Guess 1* by **clicking the response bar** on your screen. The width of the blue part indicates your guess that the group is LARGE.

You will **earn either 20 tokens or 0** by the payment rule. The payment rule is designed so that **you can secure the largest chance of earning 20 tokens by reporting your most-accurate guess**. The precise payment rule details are available by request at the end of the experiment.

Secondly, you **guess the number of other PURPLE participants who will choose P in your group** this round. Your guess is a number between 0 and 3, called ***Your Guess 2***. You choose *Your Guess 2* by clicking the responding button. And you

- **earn 20 tokens** if *Your Guess 2* is correct
- **earn 0** if *Your Guess 2* is not correct.

Summary: sequence of actions

In Phase 2, you will be in the same group as in Phase 1, and your role will also be the same.

In each of the 20 rounds in Phase 2:

1. The number of **YELLOW** participants in your group will be randomly determined: either 2 ("SMALL" group) with a 1-in-5 chance or 72 ("LARGE" group) with a 4-in-5 chance.
2. The **GRAY** participant in your group
   a. will receive a message about the size of your group.
   b. will then choose either to publicly encourage the **PURPLE** participants to choose **P** or to do nothing.
   c. will make a guess of the number of **PURPLE** participants who choose **P** if they choose to make the encouragement.
3. All **PURPLE** participants
   a. will see the **GRAY** participant's decision.
   b. will guess the percentage that the group is SMALL.
   c. will guess the number of other **PURPLE** participants who choose P.
   d. will decide to choose **P** or **Y**. All **YELLOW** participants will choose **Y**.
4. The **PURPLE** participants and **YELLOW** participants in your group will be randomly paired.
5. At the end of each round,
   a. **GRAY** participants who choose to make the encouragement **PURPLE** will know the number of **PURPLE** participants choosing **P**, and
   b. **PURPLE** participants will know the choice of the participant they are paired with.
6. The outcomes of guesses will be revealed only at the end of the experiment.

Earnings

At the end of experiment, the computer will **randomly select 5 different rounds** out of 20 rounds in Phase 2 which will be relevant to determine your payment.

If you are a **PURPLE** participant, your earnings from Phase 2 will be equal to the sum of

- **your payoffs from your *Choice* in 3 randomly selected rounds**,
- the earning from *Your Guess 1* in 1 randomly selected round, and
- the earning from *Your Guess 2* in 1 randomly selected round.

If you are a **GRAY** participant, your earnings from Phase 2 will be equal to the sum of

- **your payoffs from your *Choice* in 3 randomly selected rounds**, and
- the earning from your *Guess* in 2 randomly selected rounds.

Notice that you can NEVER be paid for both your *Choice* and your *Guess* in the same round. Earnings in tokens will be converted into Euros at the rate of: **20 tokens = 1 euro**.

5.1.2 IDEO treatment

Welcome to this study on economic decision-making.

The study consists of three phases:

in Phase 1 and phase 2 you will be matched with other participants and make some decisions.

in Phase 3 you will have to fill out a questionnaire.

At the end of the study, you will receive an amount of money which depends on your decisions and on the decisions of the other participants.

Instructions for Phase 1

Phase 1 lasts 5 rounds. In Phase 1, you will be part of a group which consists of **GRAY**, **PURPLE**, and **YELLOW** participants. **GRAY and PURPLE participants are real participants,** while **YELLOW participants are computerized**.

In each round of Phase 1, you will make a decision which may result in gaining tokens or losing tokens.

Who are the GRAY and PURPLE participants?

At the beginning of Phase 1, you will be randomly matched with 4 other real participants. One among the five of you will be randomly selected to take the role of the **GRAY** participant, while the 4 others will be **PURPLE** participants.

In all the 5 rounds of Phase 1,

- **your role will remain the same,** and
- you will be **grouped with the same other 4 real participants.**

Who are the YELLOW participants?

In your group there will also be some **computerized participants**. They will all take the role of **YELLOW** participants.

The **number** of **YELLOW** participants in your group is **randomly determined at the beginning of each round and can vary from one round to another**. This means that you will not know for sure the exact number of **YELLOW** (computerized) participants in your group. In each round, the number of **YELLOW** (computerized) participants can be

- **2** with **a 1-in-5 chance** (a 20% probability)
- **72** with **a 4-in-5 chance** (a 80% probability)

The size of your group

In each round, depending on the number of YELLOW (computerized) participants in your group, your group can be

- **"SMALL" - with 2** YELLOW participants
- **"LARGE" - with 72** YELLOW participants

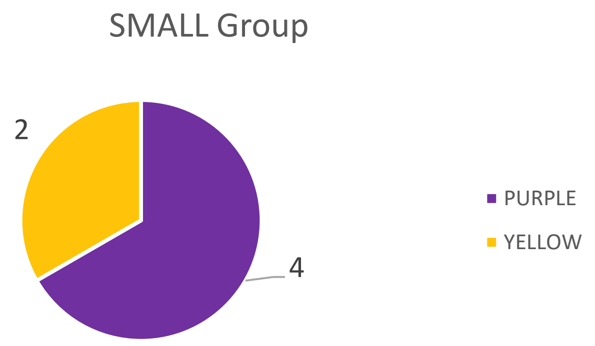


Figure 1: The composition of a "SMALL" group excluding a **GRAY** participant

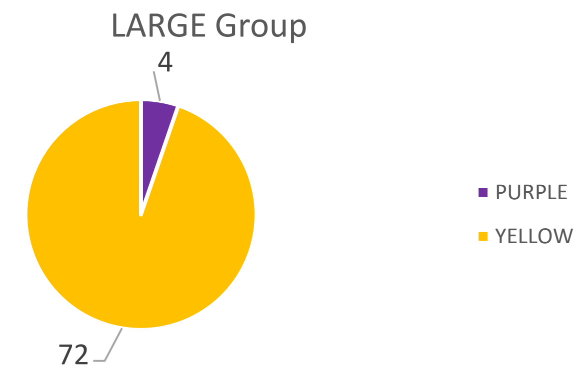


Figure 2: The composition of a "LARGE" group excluding a **GRAY** participant

Task of the PURPLE participants:

If you are a PURPLE participant, you will receive an endowment of 20 tokens at the beginning of each round.

In each round, your task is to decide whether to

- reveal your color and **choose P,**
- or do not reveal your color and **choose Y.**

After your decision, you will **be paired with another PURPLE or YELLOW (computerized)** participant in your group. This pairing is **randomly determined in each round and can vary from one round to another**. YELLOW **(computerized) participants always choose Y.**

Your payoff will depend on your choice and the choice of the participant you are paired with, as represented in the following table (in tokens).

		The other's choice	
		P	**Y**
Your choice	**P**	You **earn**: 270 The other **earns**: 270	You **lose**: 30 The other **earns**: 0
	Y	You **earn**: 0 The other **loses**: 30	You **earn**: 0 The other **earns**: 0

The table reads as follows:

1. If you choose to reveal your color, that is to **choose P**:
    - If you are paired with a participant **choosing P**, you will **earn 270 tokens**, and the other will earn 270 tokens;
    - If you are paired with a participant **choosing Y**, you will **lose 30 tokens**, and the other will earn 0 token.
2. If instead you **choose Y**, no matter who you are paired with, **you will earn 0 token.**
    - If the participant you are paired with chooses P, they will lose 30 tokens.
    - If the participant you are paired with chooses Y, they will earn 0 token.

Notice that **the chance that you are paired with a PURPLE participant** depends on the number of YELLOW (computerized) participants in your group.

If your group is **"SMALL"** (with **2** YELLOW participants),

- The chance that **you are paired with a PURPLE** participant is 3 in 5, that is **60%**
- The chance that **you are paired with a YELLOW** participant is 2 in 5, that is **40%.**

If your group is **"LARGE"** (with **72** YELLOW participants),

- The chance that **you are paired with a PURPLE** participant is 1 in 25, that is **4%**;
- The chance that **you are paired with a YELLOW** participant is 24 in 25, that is **96%.**

Task of GRAY participants:

If you are a GRAY participant, you will receive an endowment of 12 tokens at the beginning of each round.

In each round, you will be informed of your group size, either **"The Group is SMALL"** or **"The Group is LARGE"**.

After receiving the message, you need to guess

- **the number of PURPLE participants in your group that will choose P**

You will **earn**

- **20 tokens** if your guess is **correct**
- **0** if your guess is **incorrect**

Summary: sequence of actions

At the beginning of Phase 1, you will be randomly matched with 4 other real participants, and you will know your color, **GRAY** or **PURPLE**.

In each of the 5 rounds in Phase 1:

1. The number of **YELLOW** (computerized) participants in your group will be randomly determined: either 2 ("SMALL" group) with a 1-in-5 chance or 72 ("LARGE" group) with a 4-in-5 chance.
2. The **GRAY** participant in your group will receive a message about the size of your group. The **GRAY** participant will then guess how many **PURPLE** participants in your group will choose P.
3. Each **PURPLE** participant will decide to choose either **P** or **Y**. All **YELLOW** (computerized) participants will choose **Y**.
4. The **PURPLE** participants and **YELLOW** (computerized) participants in your group will be randomly paired.
5. All participants will know their payoff this round.

Earnings

At the end of experiment, the computer will **randomly select 1 of 5 rounds** in Phase 1 which will be relevant to determine your payment. Your earnings from Phase 1 will be equal to **your payoff in the randomly selected round**. Earnings will be converted into Euros at the rate of: **20 tokens = 1 euro**.

Instructions for Phase 2

Phase 2 lasts 20 rounds. In Phase 2, **you will be in the same group as in Phase 1, and your role will also be the same.** The situation you will face is also the same as in Phase 1, with a few differences, which will be described in detail in these instructions.

Tasks of the GRAY Participants

If you are a GRAY **participant,** you will receive an endowment of 12 tokens at the beginning of each round.

In each round, you will be informed of your group size, either **"The Group is SMALL"** or **"The Group is LARGE"**.

Choice

After receiving the message, you can choose:

- either to **publicly encourage the PURPLE participants to** reveal their color - **choose P**. In this way, all **PURPLE** participants in your group will see your encouragement,
- or to **do nothing**.

If you choose to **do nothing**, **your payoff will be 0**, no matter what other participants choose.

If instead you **choose to encourage PURPLE participants to choose P,** your payoff will depend on

- **your group size**,
- and on **the number of PURPLE participants choosing P in your group**.
    - **the higher** the number of **PURPLE** participants choosing P, **the higher** the payoff you will get.

The table below shows the exact payoff you would get if you choose to **encourage PURPLE participants to choose P.**

Total number of PURPLE participants choosing P	The GRAY participant's payoff if choosing to encourage PURPLE participants to choose **P**	
	"SMALL" group	"LARGE" group
0	**-4**	**-4**
1	**32**	**4**
2	**36**	**10**
3	**54**	**15**
4	**60**	**28**

This implies that, for example, if you **choose to encourage PURPLE participants to choose P,** and in your group, there are **3 PURPLE** participants choosing **P** and **1 PURPLE** participant choosing **Y,**

a. If your group is **“SMALL”**, you will **earn 54** tokens;
b. If instead, your group is **“LARGE”,** you will **earn 15** tokens.

Guess

And if you choose to encourage, you also need to make a guess of

- **the number of PURPLE participants in your group that will choose P**

You will **earn**

- **20 tokens** if your guess is **correct**
- **0** if your guess is **not correct**

Tasks of the **PURPLE** Participants

If you are a PURPLE participant, you will receive an endowment of 20 tokens at the beginning of each round.

Choice

In each round, your task is to decide whether to

- reveal your color and **choose P,**
- or do not reveal your color and **choose Y.**

The payoff structure is the same as in Phase 1.

Before making the decision, you will **first see the GRAY participant's choice,** whether they encourage all **PURPLE** participants to choose **P** or not, then you have to **make two guesses**.

Guess

Firstly, you **guess the chance that your group is LARGE in this round**. Your guess is a percentage probability from 0 to 100

- with 0 indicating 0-in-100 chance that your group is LARGE (i.e. you are sure that your group is SMALL),
- and 100 indicating a 100-in-100 chance that your group is LARGE (i.e. you are sure that your group is LARGE).

The number you provide is called ***Your Guess 1.***

You choose *Your Guess 1* by **clicking the response bar** on your screen. The width of the blue part indicates your guess that the group is LARGE.

You will **earn either 20 tokens or 0** by the payment rule. The payment rule is designed so that **you can secure the largest chance of earning 20 tokens by reporting your most-accurate guess**. The precise payment rule details are available by request at the end of the experiment.

Secondly, you **guess the number of other PURPLE participants who will choose P in your group** this round. Your guess is a number between 0 and 3, called ***Your Guess 2***. You choose *Your Guess 2* by clicking the responding button. And you

- **earn 20 tokens** if *Your Guess 2* is correct
- **earn 0** if *Your Guess 2* is not correct.

Summary: sequence of actions

In Phase 2, you will be in the same group as in Phase 1, and your role will also be the same.

In each of the 20 rounds in Phase 2:

1. The number of **YELLOW** participants in your group will be randomly determined: either 2 ("SMALL" group) with a 1-in-5 chance or 72 ("LARGE" group) with a 4-in-5 chance.
2. The **GRAY** participant in your group
   a. will receive a message about the size of your group.
   b. will then choose either to publicly encourage the **PURPLE** participants to choose **P** or to do nothing.
   c. will make a guess of the number of **PURPLE** participants who choose **P** if they choose to make the encouragement.
3. All **PURPLE** participants
   a. will see the **GRAY** participant's decision.
   b. will guess the percentage that the group is SMALL.
   c. will guess the number of other **PURPLE** participants who choose P.
   d. will decide to choose **P** or **Y**. All **YELLOW** participants will choose **Y**.
4. The **PURPLE** participants and **YELLOW** participants in your group will be randomly paired.
5. At the end of each round,
   a. **GRAY** participants who choose to make the encouragement **PURPLE** will know the number of **PURPLE** participants choosing **P**, and
   b. **PURPLE** participants will know the choice of the participant they are paired with.
6. The outcomes of guesses will be revealed only at the end of the experiment.

Earnings

At the end of experiment, the computer will **randomly select 5 different rounds** out of 20 rounds in Phase 2 which will be relevant to determine your payment.

If you are a **PURPLE** participant, your earnings from Phase 2 will be equal to the sum of

- **your payoffs from your *Choice* in 3 randomly selected rounds**,
- the earning from *Your Guess 1* in 1 randomly selected round, and
- the earning from *Your Guess 2* in 1 randomly selected round.

If you are a **GRAY** participant, your earnings from Phase 2 will be equal to the sum of

- **your payoffs from your *Choice* in 3 randomly selected rounds**, and
- the earning from your *Guess* in 2 randomly selected rounds.

Notice that you can NEVER be paid for both your *Choice* and your *Guess* in the same round. Earnings in tokens will be converted into Euros at the rate of: **20 tokens = 1 euro**.

5.1.3 HOMO treatment

Welcome to this study on economic decision-making.

The study consists of three phases:

in Phase 1 and phase 2 you will be matched with other participants and make some decisions.

in Phase 3 you will have to fill out a questionnaire.

At the end of the study, you will receive an amount of money which depends on your decisions and on the decisions of the other participants.

Instructions for Phase 1

Phase 1 lasts 5 rounds. In Phase 1, you will be part of a group which consists of **GRAY**, **PURPLE**, and **YELLOW** participants. **GRAY and PURPLE participants are real participants,** while **YELLOW participants are computerized**.

In each round of Phase 1, you will make a decision which may result in gaining tokens or losing tokens.

Who are the GRAY and PURPLE participants?

At the beginning of Phase 1, you will be randomly matched with 4 other real participants. One among the five of you will be randomly selected to take the role of the **GRAY** participant, while the 4 others will be **PURPLE** participants.

In all the 5 rounds of Phase 1,

- **your role will remain the same,** and
- you will be **grouped with the same other 4 real participants.**

Who are the YELLOW participants?

In your group there will also be some **computerized participants**. They will all take the role of **YELLOW** participants.

The **number** of **YELLOW** participants in your group is **randomly determined at the beginning of each round and can vary from one round to another**. This means that you will not know for sure the exact number of **YELLOW** (computerized) participants in your group. In each round, the number of **YELLOW** (computerized) participants can be

- **2** with **a 1-in-5 chance** (a 20% probability)
- **72** with **a 4-in-5 chance** (a 80% probability)

The size of your group

In each round, depending on the number of YELLOW (computerized) participants in your group, your group can be

- **"SMALL" - with 2** YELLOW participants
- **"LARGE" - with 72** YELLOW participants

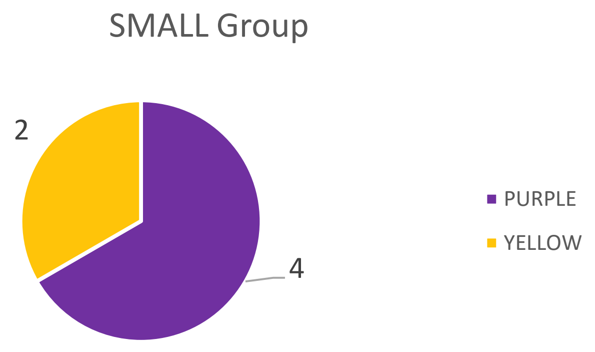


Figure 1: The composition of a "SMALL" group excluding a **GRAY** participant

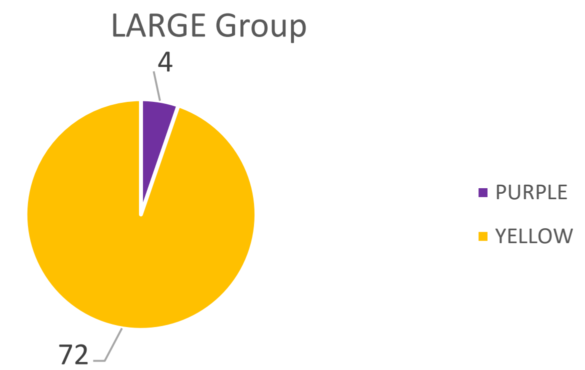


Figure 2: The composition of a "LARGE" group excluding a **GRAY** participant

Task of the PURPLE participants:

If you are a PURPLE participant, you will receive an endowment of 20 tokens at the beginning of each round.

In each round, your task is to decide whether to

- reveal your color and **choose P,**
- or do not reveal your color and **choose Y.**

After your decision, you will **be paired with another PURPLE or YELLOW (computerized)** participant in your group. The pairing rule is as follows:

- **Internal Matching** with a 1-in-20 chance (a 5% probability): All **PURPLE**s will be paired among themselves.
- **General Matching** with a 19-in-20 chance (a 95% probability): **PURPLE**s and YELLOWs will be randomly paired

This pairing is **randomly determined in each round and can vary from one round to another**. **YELLOW (computerized) participants always choose Y.**

Your payoff will depend on your choice and the choice of the participant you are paired with, as represented in the following table (in tokens).

		The other's choice	
		P	Y
Your choice	**P**	You **earn**: 270 The other **earns**: 270	You **lose**: 30 The other **earns**: 0
	Y	You **earn**: 0 The other **loses**: 30	You **earn**: 0 The other **earns**: 0

The table reads as follows:

1. If you choose to reveal your color, that is to **choose P**:
   - If you are paired with a participant **choosing P**, you will **earn 270 tokens**, and the other will earn 270 tokens;
   - If you are paired with a participant **choosing** Y, you will **lose 30 tokens**, and the other will earn 0 token.
2. If instead you **choose** Y, no matter who you are paired with, **you will earn 0 token.**
   - If the participant you are paired with chooses **P**, they will lose 30 tokens.
   - If the participant you are paired with chooses Y, they will earn 0 token.

Notice that:

If **Internal Matching** is applied, you will be paired with a **PURPLE** participant for sure.

If **General Matching** is applied, **the chance that you are paired with a PURPLE participant** depends on the number of YELLOW (computerized) participants in your group.

- If your group is **"SMALL"** (with **2** YELLOW participants),
  - The chance that **you are paired with a PURPLE** participant is 3 in 5, that is **60%**
  - The chance that **you are paired with a YELLOW** participant is 2 in 5, that is **40%.**
- If your group is **"LARGE"** (with **72** YELLOW participants),
  - The chance that **you are paired with a PURPLE** participant is 1 in 25, that is **4%**;
  - The chance that **you are paired with a YELLOW** participant is 24 in 25, that is **96%.**

Task of GRAY participants:

If you are a GRAY participant, you will receive an endowment of 12 tokens at the beginning of each round.

In each round, you will be informed of your group size, either **“The Group is SMALL”** or **“The Group is LARGE”**.

After receiving the message, you need to guess

- **the number of PURPLE participants in your group that will choose P**

You will **earn**

- **20 tokens** if your guess is **correct**
- **0** if your guess is **incorrect**

Summary: sequence of actions

At the beginning of Phase 1, you will be randomly matched with 4 other real participants, and you will know your color, **GRAY** or **PURPLE**.

In each of the 5 rounds in Phase 1:

1. The number of **YELLOW** (computerized) participants in your group will be randomly determined: either 2 (“SMALL” group) with a 1-in-5 chance or 72 (“LARGE” group) with a 4-in-5 chance.
2. The **GRAY** participant in your group will receive a message about the size of your group. The **GRAY** participant will then guess how many **PURPLE** participants in your group will choose P.
3. Each **PURPLE** participant will decide to choose either **P** or **Y**. All **YELLOW** (computerized) participants will choose **Y**.
4. The **PURPLE** participants and **YELLOW** (computerized) participants in your group will be randomly paired: either Internal Matching or General Matching.
5. All participants will know their payoff this round.

Earnings

At the end of experiment, the computer will **randomly select 1 of 5 rounds** in Phase 1 which will be relevant to determine your payment. Your earnings from Phase 1 will be equal to **your payoff in the randomly selected round**. Earnings will be converted into Euros at the rate of: **20 tokens = 1 euro**.

Instructions for Phase 2

Phase 2 lasts 20 rounds. In Phase 2, **you will be in the same group as in Phase 1, and your role will also be the same.** The situation you will face is also the same as in Phase 1, with a few differences, which will be described in detail in these instructions.

Tasks of the GRAY Participants

If you are a GRAY **participant,** you will receive an endowment of 12 tokens at the beginning of each round.

In each round, you will be informed of your group size, either **"The Group is SMALL"** or **"The Group is LARGE"**.

Choice

After receiving the message, you can choose:

- either to **publicly encourage the PURPLE participants to** reveal their color - **choose P**. In this way, all **PURPLE** participants in your group will see your encouragement,
- or to **do nothing**.

If you choose to **do nothing**, **your payoff will be 0**, no matter what other participants choose.

If instead you **choose to encourage PURPLE participants to choose P,** your payoff will depend on

- **your group size**,
- and on **the number of PURPLE participants choosing P in your group**.
  - **the higher** the number of **PURPLE** participants choosing P, **the higher** the payoff you will get.

The table below shows the exact payoff you would get if you choose to **encourage PURPLE participants to choose P.**

Total number of **PURPLE** participants choosing P	The GRAY participant's payoff if choosing to encourage **PURPLE** participants to choose **P**	
	"SMALL" group	"LARGE" group
0	**-4**	**-4**
1	**32**	**4**
2	**36**	**10**
3	**54**	**15**
4	**60**	**28**

This implies that, for example, if you **choose to encourage PURPLE participants to choose P,** and in your group, there are **3 PURPLE** participants choosing **P** and **1 PURPLE** participant choosing **Y,**

a. If your group is **"SMALL"**, you will **earn 54** tokens;
b. If instead, your group is **"LARGE",** you will **earn 15** tokens.

Guess

And if you choose to encourage, you also need to make a guess of

- **the number of PURPLE participants in your group that will choose P**

You will **earn**

- **20 tokens** if your guess is **correct**
- **0** if your guess is **not correct**

Tasks of the **PURPLE** Participants

If you are a PURPLE participant, you will receive an endowment of 20 tokens at the beginning of each round.

Choice

In each round, your task is to decide whether to

- reveal your color and **choose P,**
- or do not reveal your color and **choose Y.**

The payoff structure is the same as in Phase 1.

Before making the decision, you will **first see the GRAY participant's choice,** whether they encourage all **PURPLE** participants to choose **P** or not, then you have to **make two guesses**.

Guess

Firstly, you **guess the chance that your group is LARGE in this round**. Your guess is a percentage probability from 0 to 100

- with 0 indicating 0-in-100 chance that your group is LARGE (i.e. you are sure that your group is SMALL),
- and 100 indicating a 100-in-100 chance that your group is LARGE (i.e. you are sure that your group is LARGE).

The number you provide is called ***Your Guess 1.***

You choose *Your Guess 1* by **clicking the response bar** on your screen. The width of the blue part indicates your guess that the group is LARGE.

You will **earn either 20 tokens or 0** by the payment rule. The payment rule is designed so that **you can secure the largest chance of earning 20 tokens by reporting your most-accurate guess**. The precise payment rule details are available by request at the end of the experiment.

Secondly, you **guess the number of other PURPLE participants who will choose P in your group** this round. Your guess is a number between 0 and 3, called ***Your Guess 2***. You choose *Your Guess 2* by clicking the responding button. And you

- **earn 20 tokens** if *Your Guess 2* is correct
- **earn 0** if *Your Guess 2* is not correct.

Summary: sequence of actions

In Phase 2, you will be in the same group as in Phase 1, and your role will also be the same.

In each of the 20 rounds in Phase 2:

1. The number of **YELLOW** participants in your group will be randomly determined: either 2 ("SMALL" group) with a 1-in-5 chance or 72 ("LARGE" group) with a 4-in-5 chance.
2. The **GRAY** participant in your group
   a. will receive a message about the size of your group.
   b. will then choose either to publicly encourage the **PURPLE** participants to choose **P** or to do nothing.
   c. will make a guess of the number of **PURPLE** participants who choose **P** if they choose to make the encouragement.
3. All **PURPLE** participants
   a. will see the **GRAY** participant's decision.
   b. will guess the percentage that the group is SMALL.
   c. will guess the number of other **PURPLE** participants who choose P.
   d. will decide to choose **P** or **Y**. All **YELLOW** participants will choose **Y**.
4. The **PURPLE** participants and **YELLOW** (computerized) participants in your group will be randomly paired: either Internal Matching or General Matching.
5. At the end of each round,
   a. **GRAY** participants who choose to make the encouragement **PURPLE** will know the number of **PURPLE** participants choosing **P**, and
   b. **PURPLE** participants will know the choice of the participant they are paired with.
6. The outcomes of guesses will be revealed only at the end of the experiment.

Earnings

At the end of experiment, the computer will **randomly select 5 different rounds** out of 20 rounds in Phase 2 which will be relevant to determine your payment.

If you are a **PURPLE** participant, your earnings from Phase 2 will be equal to the sum of

- **your payoffs from your *Choice* in 3 randomly selected rounds**,
- the earning from *Your Guess 1* in 1 randomly selected round, and
- the earning from *Your Guess 2* in 1 randomly selected round.

If you are a **GRAY** participant, your earnings from Phase 2 will be equal to the sum of

- **your payoffs from your *Choice* in 3 randomly selected rounds**, and
- the earning from your *Guess* in 2 randomly selected rounds.

Notice that you can NEVER be paid for both your *Choice* and your *Guess* in the same round. Earnings in tokens will be converted into Euros at the rate of: **20 tokens = 1 euro**.

5.2 Questionnaire

After completing two phases of the experiment, each participant will fill out a questionnaire. The questionnaire is the same in all treatments. First, the participant will answer the following set of unincentivized background questions, including sex, age, birthplace, prior participation in experiments, and previous experience with game theory courses. Then, the participant will respond to several questions eliciting their risk preferences, social preferences on lying, prosocial behaviors, and altruism.

Risk preferences was be elicited using an incentivized Bomb Risk Elicitation Task (Crosetto and Filippin, 2013). The exact text used in the task is provided below.

6 *On the right side, you can see a field composed of 5*5=25 boxes. Behind one of these boxes a bomb is hidden; the remaining 24 boxes are empty. You do not know where the bomb is. You only know that it can be in any place with equal probability. Your task is to choose how many boxes to collect. Boxes will be collected in numerical order, so you will be asked to choose a number between 1 and 25.*

> *If you happen to have collected the box in which the time bomb is located – i.e., if your chosen number is greater than, or equal to, the number of the box with bomb*
>
> *- you will earn zero.*
>
> *If the time bomb is located in a box that you did not collect – i.e., if your chosen number is smaller than the number of the box with bomb*
>
> *- you will earn an amount in tokens equivalent to the number you have chosen.*

The preference regarding lying behaviors will be elicited using a non-incentivized survey question (Schudy et al., 2024):

Please imagine the host of a live radio show calls you to participate in a radio raffle. Your task is the following: you must flip a coin 4 times and report the number of "tails" you flipped. You will receive 10 pounds for each reported "tail". You know that there is no way the host can verify whether you reported the true number of "tails" you flipped.

7 *Situation 1: Imagine the unlikely case occurs that you flip 4 tails. How many tails will you report to the host? [Choice 0/1/2/3/4]*

8 *Situation 2: Imagine the unlikely case occurs that you flip 0 tails. How many tails will you report to the host? [Choice 0/1/2/3/4]*

The preference regarding prosocial behaviors and altruism will be elicited using several non-incentivized survey questions (Falk et al., 2018):

9 *Imagine the following situation: you are shopping in an unfamiliar city and realize you lost your way. You ask a stranger for directions. The stranger offers to take you with their car to your destination. The ride takes about 20 minutes and costs the stranger about 20 Euro in total. The stranger does not want money for it. You carry six bottles of wine with you. The cheapest bottle costs 5 Euro, the most expensive one 30 Euro. You decide to give one of the bottles to the stranger as a thank-you gift. Which bottle do you give? Respondents can choose from the following options: The bottle for 5, 10, 15, 20, 25, or 30 Euro)*

10 *Imagine the following situation: you won 1,000 Euro in a lottery. Considering your current situation, how much would you donate to good cause? (Values between 0 and 1000 are allowed)*

11 *How do you assess your willingness to share with others without expecting anything in return when it comes to good cause? Please use a scale from 0 to 10, where 0 means you are completely unwilling to share and a 10 means you are very willing to share. You can also use the values in between to indicate where you fall on the scale.*