Chlaß, Nadine; Miettinen, Topi

Working Paper

Commitment in sequential bargaining: An experiment

Helsinki GSE Discussion Papers, No. 42

Provided in Cooperation with:
Helsinki Graduate School of Economics

Suggested Citation: Chlaß, Nadine; Miettinen, Topi (2025) : Commitment in sequential bargaining: An experiment, Helsinki GSE Discussion Papers, No. 42, ISBN 978-952-7543-41-2, Helsinki Graduate School of Economics, Helsinki

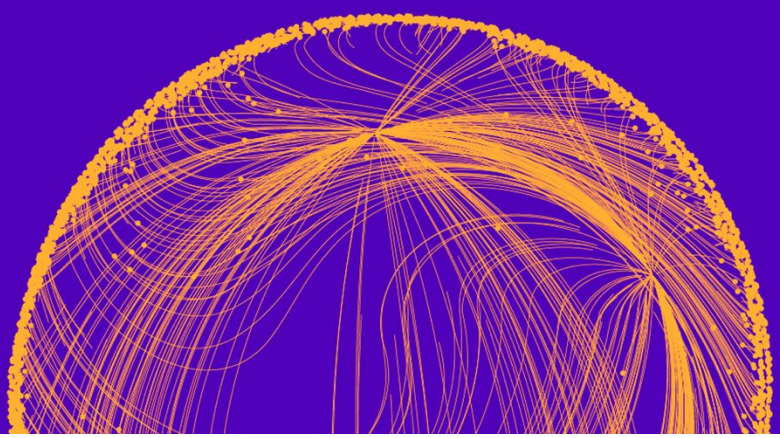
This Version is available at:
https://hdl.handle.net/10419/330261

Helsinki GSE

HELSINKI GSE DISCUSSION PAPERS 42 · 2025

Commitment in Sequential Bargaining – An Experiment

Nadine Chlaß
Topi Miettinen



HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

HANKEN

Aalto University

Helsinki GSE Discussion Papers

Helsinki GSE Discussion Papers 42 · 2025

Nadine Chlaß and Topi Miettinen
Commitment in Sequential Bargaining - An Experiment

ISBN 978-952-7543-41-2 (PDF)
ISSN 2954-1492

Helsinki GSE Discussion Papers:
https://www.helsinkigse.fi/discussion-papers

Helsinki Graduate School of Economics
PO BOX 21210
FI-00076 AALTO
FINLAND

Helsinki, August 2025

Commitment in Sequential Bargaining - An Experiment*

Nadine Chlaß †& Topi Miettinen‡

August 16, 2025

Abstract

Schelling (1956) first clarified how power to reduce one's freedom of choice might benefit a bargaining party. A commitment to reject proposals, when successful, may force concessions from opponents who otherwise might have an upper hand. This paper experimentally studies credible commitments prior to a sequential ultimatum bargaining game. We find that pre-emptive commitment strategies are exploited by the responders but less than predicted by theory. In a game where a responder can unilaterally precommit, she faces the same incentives as a proposer in an ultimatum game. Yet, the observed responder commitments are less aggressive than proposals by proposers in the ultimatum game. In a simultaneous commitment game, proposers who cannot benefit from committing are nevertheless observed to commit. The observed within-treatment payoff-differences between the two parties do not comply with the theoretical predictions in the commitment variants of the game. Surprisingly in late rounds, allowing for pre-commitment yields almost 100% efficiency both when only responders and when also the proposers are allowed to commit although the ultimatum game features significant inefficiencies even in late rounds. We discuss four complementary behavioral explanations and find that reciprocity and concern for equality of opportunity are consistent with the observed patterns. Empirically, we observe that ethical criteria underlying preferences for equal opportunity are at work.

JEL: D74, C91, D02, D63, D91

Keywords: bargaining, precommitment, reciprocity, equality of opportunity, moral judgment

1 Introduction

Schelling (1956) argued that credible and well-communicated commitments in bargaining can be beneficial by forcing concessions from others. However, if multiple parties act simultaneously (or without knowledge of each other's moves) making commitments with stochastic success, that can lead to incompatible bargaining positions. If revoking the commitments is costly or impossible, such strategic posturing can result in inefficient delay or impasse. Schelling's perspective stands in stark contrast to Nash (1953) who motivated his efficient rationality prescription in bargaining by, likewise, using a simultaneous demands game to which some exogenous uncertainty was introduced.

Crawford (1982); Muthoo (1992, 1996) and Ellingsen and Miettinen (2008), among others, developed formalizations of the original insights showing how central the assumptions about the revoking of commitments are to the efficiency and distributive implications of commitment. In particular, when revoking is strategic and non-stochastic

*The authors equally share first authorship. We thank Yrjö Jahnsson Foundation for financial support, and Andrés Perea for valuable comments and support. All errors are our own.

†Friedrich-Schiller University Jena, Carl-Zeiss-Str. 3, 07743 Jena, Germany. Email: nadine.chlass@uni-jena.de.

‡Hanken School of Economics & Helsinki GSE, Arkadiankatu 22, 00100 Helsinki, Finland. Email: topi.miettinen@hanken.fi

and costs are complete information, outcome is efficient and commitment can typically only have distributive effects (Muthoo, 1992, 1996; Miettinen and Perea, 2015).[1]

In this study, we experimentally investigate the distributive and inefficiency effects of access to precommitment, when backing down is prohibitively costly or impossible. The underlying negotiation is an *ultimatum* game in strategic form (with a monotone responder strategy) where a proposer makes a take-it-or-leave-it (TIOLI) offer, and a responder simultaneously chooses a minimum acceptable offer (MAO).[2] This baseline treatment is extended by allowing for precommitment by the responder only (*unilateral* or responder commitment) or simultaneously by both the proposer and the responder (*simultaneous* commitment). The unique prediction in *ultimatum* is well known when agents are self-interested and sequentially rational: the proposer holds the initiative and can reap all gains from trade as the responder is always better off accepting even a tiny share of the pie.

However, in *unilateral*, the strategic incentives are reversed: the responder should commit to a position which leaves the proposer at the impasse payoff. Thus, the responder should receive all the gains from trade.[3] Indeed, the incentives of the proposer when proposing in the ultimatum game are isomorphic to those of the responder when precommitting in the unilateral commitment game.

The predictions for *simultaneous* are more subtle. The first stage of the game mimics the Nash (1953) demand game and, indeed, any perfectly compatible pair of demands is a Nash equilibrium of the game. As Schelling (1956) verbally argues, however, any genuine strategic uncertainty about the simultaneous choice of the opponent induces a severe risk of impasse.[4] As he tacitly argues, a descriptive model should take this uncertainty seriously. In our setup, it is the proposer who has the greater risk of losing her lion's share of the pie in the ensuing ultimatum game. In fact, the proposer has nothing to gain by precommitting, as she will yield the same or a higher payoff by not committing and allocating the residual gains from trade to herself in the proposal of the ensuing ultimatum game. Thus, if we agree with the logic of the refinement, the proposer should not commit, and in response, the responder should make the same commitment as in *unilateral*.

The patterns which we observe in *ultimatum* look standard: proposers take a lion's share of the pie and inefficient rejections of small proposals by responders are frequent.[5] Yet, we see that the responders commit approximately to the 50-50 split and little more in *unilateral*, and that their commitments are significantly less aggressive than the proposals of the proposers in *ultimatum*. This is contrary to the prediction of the lion's share now going to the responders. When also proposers are granted commitment power and they can exercise it simultaneously with the responders, we see that responders commit even more tightly around the 50-50 split. Proposers do not refrain from committing, as predicted by the refinement, but rather also commit to the 50-50 split which is in line with behavior in a focal Nash equilibrium. Moreover, efficiency is much higher in both commitment games than in *ultimatum*, at least after some learning and adaptation in a repeated perfect strangers matching. Thus, contrary to the theoretical prediction, we find both efficiency and redistributive implications of access to precommitment.

To account for these patterns, we consider four complementary explanations: (i) fairness and inequality aversion (Bolton, 1991; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000), (2) quantal-response equilibrium (McKelvey and Palfrey, 1998; Yi, 2005), (3) (tractable model of) reciprocity (Cox et al., 2007, 2008), and (4) concern for the equality of opportunity (Chlaß et al., 2019). While all possibly contribute to the patterns, we argue that the

[1] Renou (2009) and Bade et al. (2009) show very generally that, when the initial precommitments are certain to succeed, without cost, and last forever, any equilibrium of the underlying game B is still an equilibrium outcome of the game with commitments. This holds for any efficient bargaining game, in particular. See Kambe (1999a) for a reputational model with asymmetric information about commitment power. See also Li (2011); Britz (2013); Ellingsen and Miettinen (2014); Chung and Wood (2019).

[2] Notice that choosing a minimum acceptable offer is equivalent to choosing a monotonically weakly increasing acceptance strategy.

[3] See (Miettinen and Perea, 2015) for a detailed argument in an epistemic model.

[4] The notion of trembling-hand perfection to capture implications of genuine, even if small, strategic uncertainty on predictions has been introduced by Selten (1975). An example of an epistemic concept applicable to extensive form games is Asheim and Perea (2005)

[5] It is also well known that small proposals are rejected by responders, that observed proposals to the responder average at 40-50%, and that proposals at 20% and below are very likely to be rejected (Güth and Kocher, 2014).

last two are most consistent with the overall evidence.

Despite a voluminous theoretical literature, there are few experimental studies of complete information precommitment in bargaining.[6] In an early exception, Fershtman and Gneezy (2001) study ultimatum bargaining where commitment arises by mandated delegates.[7] The study finds that the use of delegates, whose incentives are observable to the other side, benefits the party sending the delegates whether sent by the proposer or the responder. Their study, however, does not allow for risk of impasse through simultaneous decisions; in our experiment, this is present in the simultaneous commitments treatment.

Since in our commitment treatments, players are given an opportunity to limit the choice sets of each other, our study is also relates to the seminal study of Falk and Kosfeld (2006) on hidden costs of control in a principal-agent setting. The original contribution illustrated that the observed limitation of the choice set and the lower implied effort by the agent cannot be explained by reciprocity theories where payoffs explicitly depend on beliefs (Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006).[8] In the present paper we show, however, that a tractable model of reciprocity (Cox et al., 2007, 2008) can organize our behavioral patterns in all three treatments. Our bargaining interaction is richer, however, as from a theoretical perspective, there are three stages and sometimes simultaneous moves, and both parties are still active once commitment(s), which induce control as defined by Falk and Kosfeld (2006), have been established.

Another closely related study is Chen et al. (2024). The authors consider a setup where responders choose both a minimum acceptable offer (MAO) and a cost of backing down (CBD) from that commitment, both of which are communicated to the proposer before a proposal is made. Choosing a CBD lower than a MAO constitutes a partial commitment, where it is optimal to revoke proposals equal to the MAO. When the CBD is sufficiently high vis-à-vis the MAO, revoking is prohibitively costly and credible commitment is established. The central research questions ask whether commitments are partial or credible and why partial commitment is commonly observed, as opposed to credible commitment. In the study, proposers concede more to credible commitments. The proposer could also choose to commit, and the experiment varies whether or not the proposer's commitment becomes known to the responder prior to committing. The authors find that responder commitment behavior is not responsive to information about proposer commitment. Thus, the risk of impasse does not influence bargaining behavior.[9] Note that in our study, commitment is always credible by design and revoking is impossible. We exogenously vary players' access to precommitment and find that proposers' access to commitment causally affects responder commitments.

A more remotely related line of literature is interested in the variation in threat power of the responder (Güth and Huck, 1997; Fellner and Güth, 2003; Rodriguez-Lara, 2016). Rather than studying the responder's capacity to express threats before the proposal is formed, the contributions are interested in costs that a rejection inflicts on the two parties. There is also an experimental literature studying endogenous timing in duopoly games (Huck et al., 2002; Fonseca et al., 2006; Müller, 2006) where early timing of capacity choice can be interpreted as a form of commitment.

The paper is organized as follows. In section 2, we derive the key theory underlying our predictions which are listed in Section 2.4. Section 3 presents the experimental design. Section 4 covers the results. We discuss the evidence and potential explanations in Section 5. Section 6 concludes.

[6] Embrey et al. (2015); Fanning and Kloosterman (2022); Heggedal et al. (2022) investigate reputational bargaining (Myerson, 1991; Kambe, 1999b; Abreu and Gul, 2000) in the laboratory, building on the idea that even a small chance of existence of crazy or stubborn types can be used to build commitment power.

[7] See for instance Schelling (1960) and Fershtman and Judd (1987).

[8] Von Siemens (2013) later showed that the model of Dufwenberg and Kirchsteiger (2004), to which incomplete information about types is introduced, can also explain the patterns of Falk and Kosfeld (2006) for specific parameter values. Other asymmetric information explanations, which are challenging to apply to our rich three-stage interactions, include Sliwka (2007) and Ellingsen and Johannesson (2008).

[9] Swope et al. (2014) study precommitment in a multilateral context where discount factors vary.

2 Theory

2.1 Model

There are two players, A and B, who must reach an agreement about the division of 12 euros. The set of possible divisions is given by

$$D := \{(x_A, x_B) : x_i \in \mathbb{N} \text{ and } x_A + x_B = 12\}.$$

where x_i is the euro amount assigned to Player i in the agreement.

Players A and B use the following bargaining procedure. At the beginning, either nobody (*ultimatum*), B only (*unilateral*), or A and B simultaneously (*simultaneous*) choose *commitment levels* $c_A, c_B \in \{0, ..., 12\}$. The commitment of a player is automatically set to zero if she is not active in the commitment stage, i.e. for both players in *ultimatum* and for Player A in *unilateral*. The commitment levels become known to both players, and they constitute credible threats in the sense that sharings where the player gets less than her threat will not be implemented and rather an impasse payoff, $(2,0)$, is realized. Once the threats are known, Player A proposes a division $(x_A, x_B) \in D$ under the condition that only proposals with $x_A \geq c_A$ can be implemented; $x_A < c_A$ automatically leads to impasse payoffs. Player B decides whether to accept or reject the proposal under the condition that she can only accept offers with $x_B \geq c_B$. If she accepts, (x_A, x_B) is the final outcome. If she rejects, the impasse payoff $(2,0)$ is realized. Notice that Player A can guarantee a payoff of 2 by committing to $c_A \leq 2$ and then proposing $(2,10)$. Thus, *gains from trade* are $12 - 2 = 10$.

Within our bargaining procedure above, the interpretation of the commitment levels is thus that the proposer commits not to make a serious offer less than her commitment level to herself, whereas the responder commits to reject any offer that would give her less than her commitment level.

2.2 Treatments

There were three alternative treatments: first, the standard *ultimatum* game played in strategic form imposing a monotonically increasing responder strategy (Thaler, 1988). In *ultimatum*, the responder does not have access to an observable precommitment; second, the *unilateral* commitment game where, prior to the ultimatum game, only Player B can commit, and this commitment is then observed by the proposer before making the proposal; and third, the *simultaneous* commitment game where both Player A and Player B can precommit. Both commitments are then observed by both players before a proposal and a (revised) minimum acceptable offer are made by the proposer and the responder, respectively.

In the *ultimatum* treatment, Player A chooses a division of 12 euros: a share in whole euros for Player B, x_B, and the residual for herself, $12 - x_B$. Simultaneously, Player B chooses the least share she is willing to accept (minimal acceptable offer, MAO), m_B. If $x_B \geq m_B$, the compensation of the round is x_B for B and $12 - x_B$ for A; otherwise 0 and 2, respectively.

In the *unilateral* treatment, Player B first chooses her commitment, an integer euro amount c_B, while the commitment of A is automatically set to zero, $c_A = 0$. Once Player B has chosen her commitment, this latter is then communicated (on computer screens) to both players. Thereafter, the ultimatum game described above is played with the obvious additional constraint that the sharing of A is implemented only if $x_B \geq max\{c_B, m_B\}$. If $max\{c_B, m_B\} > x_B$, payoffs are 0 and 2 for Players B and A, respectively.

In the *simultaneous* treatment, Player A and Player B first simultaneously choose c_A and c_B. These constitute the commitments which are communicated to both players. Thereafter, the ultimatum game described above is played, and each player receives the proposed monetary share if and only if $x_B \geq max\{c_B, m_B\}$ and $x_A \geq c_A$.

Otherwise, the payoffs are 2 and 0 for Players A and B, respectively.

An obvious challenge in this context was to keep the instructions for all three treatments as similar as possible such that potential treatment effects would not be caused by differences in readability and complexity. We therefore framed the ultimatum game such that the responder sets her MAO in a Stage 1. Since she reports her decision to accept or reject by means of the strategy method (i.e. a MAO), she still conditions her choice on the proposer's proposal. In Stage 2, the proposer makes her proposal without knowing the MAO chosen by the responder in Stage 1. Notice that since the responder's MAO is not observable to the proposer in *ultimatum*, the strategic incentives are unaffected by this timing. The self-interested sequentially rational (or more narrowly self-interested subgame perfect equilibrium) solution still predicts that the proposer receives the entire pie.

Treatment *unilateral* differs from *ultimatum* in that the MAO chosen by the responder is revealed to the proposer before forming the proposal. The MAO thus becomes a pre-emptive commitment granting the responder the first-mover advantage. To preserve the ultimatum game structure in the underlying game, the responder can also revise her MAO in Stage 2.[10]. It is important to notice that, as already mentioned in the introduction, the results for *ultimatum* are completely standard.

2.3 Analysis

In line with the literature, we assume self-interest and apply the subgame-perfect equilibrium (SPE) to derive the benchmark predictions in most of the analysis.

Throughout and for simplicity, we adopt a tie-breaking rule and assume that an indifferent player chooses the more efficient action. In particular, an indifferent responder accepts an efficient proposal.

In addition, we discuss some alternatives — implications of either weakening SPE to Nash equilibrium or strengthening it by using some further refinements. We also discuss the implications of relaxing self-interest and consider various models of other-regarding concerns in Section 2.4, 5, and 6. Therein, we also discuss how the adoption of the logit quantal-response equilibrium would influence the predictions.

Applying backward induction first to the acceptance-rejection decision and then to the proposals made yields a unique prediction about the outcome given a commitment profile (c_A, c_B): the proposer proposes $x_B = c_B$ if $c_B \leq 12 - max\{2, c_A\}$ resulting in payoffs $(12 - x_B, x_B)$; if $max\{2, c_A\} > 12 - c_B$, the implied payoffs are $(2, 0)$. Let us next use this reasoning to derive the predictions in our three treatments.

Ultimatum game. In the ultimatum game, the commitments are automatically set to zero. Thus, the prediction is the standard $(12, 0)$ where all gains from trade accrue to the proposer who has a first-mover advantage.[11]

Unilateral commitment game. Player A's commitment is automatically set to zero, $c_A = 0$. Given commitment c_B, the proposal of Player A is $(12 - c_B, c_B)$ if $c_B \leq 10$. If $c_B > 10$, the proposer makes a proposal which will be rejected. The payoffs are

$$\begin{cases} (12 - c_B, c_B) & \text{if } c_B \leq 10 \\ (2, 0) & \text{if } c_B > 10. \end{cases}$$

Notice that Player B's payoff is strictly increasing in c_B up to 10. Thus, Player B optimally chooses commitment $c_B = 10$. Observing the commitment by B, Player A proposes $(2, 10)$ and Player B accepts.

Simultaneous commitment game. The simultaneous commitment game is essentially a Nash demand game which is followed by an ultimatum game if the demands (simultaneous commitments) are less than compatible.

[10] We would have obtained completely symmetric instructions by eliciting responders' MAOs a second time in Stage 2 of *ultimatum* but were concerned about inducing demand effects by asking for the same decision twice in that responders would have felt *induced* to revise.

[11] Since each pie-share must be an integer, there is another self-interested SPE with sharing (11,1) where the responder rejects the proposal (12,0). However, the refinement requires an indifferent responder to accept an offer which rules out this SPE.

The residual pie is shared by the ultimatum game. In the game concerning the residual pie, Player A is known to reap all the residual gains from trade. The payoffs in the ensuing game are thus

$$\begin{cases} (12-c_B, c_B) & \text{if } max\{2, c_A\} + c_B \leq 12 \\ (2,0) & \text{if } max\{2, c_A\} + c_B > 12. \end{cases}$$

Every just compatible pair of commitments where $c_A \geq 2$ can be supported as a (self-interested) Nash equilibrium of (the first stage of) *simultaneous* (Nash, 1953). Each of these just compatible pairs of commitments can also be supported as a SPE of the game: assume that in every subgame following the commitment stage of *simultaneous*, proposer proposer $(12-c_B, c_B)$ irrespective of the commitment of Player A. It is easy to see that this strategy is part of an SPE and result in payoffs $(12-c_B, c_B)$ when $c_A \leq 12-c_B$ and payoffs $(2,0)$ when $c_A > 12-c_B$. Then, in the commitment stage, a best response by A[12] to any $c_B \leq 10$ is $12-c_B$ and the best-reponse by B to any $c_A \geq 2$, including $c_A = 12-c_B$, is $12-c_A$. Thus, $12-c_B$ and c_B constitute a Nash equilibrium of the game.

Although it is important to remember that any pair of just compatible demands can be sustained as an SPE in *simultaneous*, we will here use an elimination argument to refine the SPE outcomes to single out a prediction of interest. It is generally known that introducing any stochasticity to the game, e.g. about the pie size or players' actions, may drastically reduce the set of equilibria.[13] In this particular game, if Player A faces even small uncertainty *about the commitment of Player B*, she can wait and optimally match the proposal to the realized c_B in the ultimatum game[14]: to see this, consider just compatible commitments $c_A^* + c_B^* = 12$ and a belief of Player A that places positive probability on commitment $c_B > c_B^*$, then deviating to $c_A \leq 2$ and making a proposal $(12-c_B, c_B)$ in the ultimatum game. This yields a strictly higher payoff than committing to c_A^*. More generally, any commitment strategy $c_A > 2$ is weakly dominated by $c_A \leq 2$ and making a proposal $(12-c_B, c_B)$. Player A can only suffer from making a commitment whatever the commitment of Player B.[15]

Let us now apply this refinement to *simultaneous*. We eliminate the weakly dominated commitments of Player A in the simultaneous commitment game. The resulting Player B payoff is strictly increasing in c_B up to 10. Player B therefore optimally chooses commitment level $c_B = 12-2 = 10$, and the outcome coincides with that *unilateral*.[16]

Remarks. The proposer faces a first-mover *disadvantage,* rather than a first-mover advantage in the *simultaneous* commitment game. Recall that $(2,0)$ is the outcome if the proposal is rejected. So, Player A, the proposer, gets the minimal amount she would still accept, whereas Player B, the responder, gets all the gains from trade. Thus, the proposer gets exactly what she would obtain as a responder in the procedure without commitment. So, once we consider the refinement to *simultaneous*, introducing commitment perfectly reverses the outcome (Miettinen and Perea, 2015). In this sense, the commitment patterns of Player B should look like a mirror image of proposer behavior in the ultimatum game. If we do not introduce the refinement in *simultaneous*, then the outcome predicted by self-interested SPE in *simultaneous* is any $(12-x_B, x_B)$ with $0 \leq x_B \leq 10$. In the other two games, the

[12] Another best-reponse is not to commit and then propose $(12-c_B, c_B)$.

[13] Nash (1953) introduced a perturbation to the pie and concluded with a unique efficient Nash equilibrium. Güth et al. (2004) introduced an option to wait to Nash's setting and concluded with two pure strategy equilibria where one player received almost all surplus. Ellingsen and Miettinen (2008) introduced stochastic success of commitments and an option to wait and concluded with a unique equilibrium where both attempt commitment to the entire pie.

[14] This insight can be formalized as a trembling hand perfect equilibrium (Selten, 1975). It also features in the idea of eliminating weakly dominated actions.

[15] Since in any QRE, B's commitment has full support and is stochastic, this argument also shows why just compatible commitments in *simultaneous* cannot constitute a QRE of the game. We will return to this issue in the following subsection.

[16] The epistemic foundations are provided by Asheim and Perea (2005). Miettinen and Perea (2015) analyze epistemic foundations and derive general solutions in the context of costly precommitment in sequential bargaining.

SPE-prediction is unaffected by the refinement: $(12,0)$ in *ultimatum* and $(2,10)$ in *unilateral*.

2.4 Hypotheses

Our central research questions relate to the optimal exploitation of threats:

1. Whether credible threats, i.e. commitments, are used and whether the payoffs of the players reflect these threats. Are the commitments used optimally?

2. Do Player B commitments in commitment games mirror Player A proposals in the ultimatum game?

3. Does access to commitments result in redistribution? Does it result in inefficiency?

The (refined) theory predicts that A does not benefit from and, thus, does not use precommitment, whereas B uses commitment opportunities. B's commitment as a share of gains from trade in the commitment games should look like a mirror image of Player A proposals in the ultimatum game (as a share of gains from trade). Outcomes should be efficient. Throughout, we make the assumption that an indifferent player chooses the more efficient action. In particular, an indifferent responder accepts an efficient proposal. Based on the theory presented in Section 2, we have the following predictions.[17]

1. ***Exploitation of commitments.*** *In unilateral and simultanous, Player B commits to all gains from trade (10). (Player A commits to her impasse payoff 2 or lower in the simultaneous commitments game.) Player A then proposes 10 to Player B and 2 to herself. Player B accepts all offers equal to or greater than 10.*[18] *In ultimatum, Player A proposes 12 (= her impasse payoff 2 + all gains from trade, 10) to herself and 0 to B. Player B accepts all offers.*[19] Thus, all gains from trade accrue to A in ultimatum and to B in the commitment games.

2. ***Reversal of strategic advantage.*** *The distribution of proposals x_A of Player A in ultimatum coincides with the distribution of c_B+2 of Player B in the commitment games (the proposed distribution of gains from trade coincides).*

3. ***Redistribution and efficiency.*** *Player B payoffs are higher in the ultimatum than the commitment games than in . Player A's payoffs coincide in the two commitment games. Player A's payoff is higher (lower) in ultimatum (in commitment) games than that of Player B. All outcomes are efficient. Thus, there are no differences in efficiency between the treatments.*

A few further remarks about the predictions are warranted.

Nash equilibrium. In all three games, any efficient distribution of gains from trade can be supported by self-interested Nash equilibrium play. Thus, Nash equilibrium predictions do not differ between the games. In any Nash equilibrium, Player A receives at least 2, and Player B receives at most 10. In the simultaneous commitment game, the usual inefficient Nash equilibrium also exists where players commit to demand the entire pie, resulting in payoffs two and zero. Notice that many Nash equilibria are supported by non-credible play off the Nash equilibrium path. Similar inefficient Nash equilibria also exist in the other two games.[20]

[17] An epistemic concept which yields these predictions is sequential and quasi-perfect (self-interested) rationalizability (Asheim and Perea, 2005) and an equilibrium notion is (self-interested) extensive form trembling hand perfection (Selten, 1975).

[18] Due to a discrete set of possible divisions, there exists another equilibrium in which B commits to 9 and A proposes 9 to B.

[19] In another equilibrium, A proposes 11 for herself and 1 for B. In the latter case, B rejects zero and accepts all other offers.

[20] That any efficient division can be supported as a Nash equilibrium is discussed in Gale et al. (1995); Yi (2005). It is easy to see how to extend the argument to the commitment games. The tie-breaking we apply to SPE that a player should choose a more efficient strategy if indifferent, would rule out the inefficient Nash equilibria if applied to refine Nash equilibria.

Fairness ideals. Since fairness ideals are known to have explanatory power in bargaining, two alternative solutions are worth taking note of: (i) splitting the gains from trade equally, *split-the-difference*, would result in payoffs (7,5); (ii) splitting the total payoff equally would result in payoffs (6,6) (Dufwenberg et al., 2017; Binmore et al., 1998).

Quantal response equilibrium (QRE). In extensive form QRE (McKelvey and Palfrey, 1998), each action is chosen with positive probability at each player node. Actions with higher payoffs, given equilibrium choices of others, are chosen with higher probability. In QRE, smaller proposals reduce the payoff difference between accepting and rejecting. Therefore, smaller proposals are accepted with a lower probability. This implies that small shares are also proposed with lower probability, and unilateral commitments by Player B close to 10 are less likely than less ambitious commitments. Finally, in the simultaneous move commitment game, commitments by Player B stochastically dominate commitments by Player A. This is due to Player A's strategic uncertainty about Player B's commitment and, thus, strictly higher payoff to not committing and rather matching the ensuing proposal with B's actual commitment. **Tractable models of reciprocity.** According to tractable models of reciprocity (Cox et al., 2007, 2008), refraining from committing, in our bargaining games, is a kind act since it does not rule out high-payoff opportunities from the opponent in the sequel. Likewise, greater proposals by the proposer preserve high-payoff opportunities for the responder. In these reciprocity models, altruistic concerns are promoted by such kind acts. Moreover, actions which rule out equal payoff opportunities trigger negative reciprocation and spite in the sequel. Thus, by refraining from making ambitious commitments above the equal split, the responder limits negative reciprocation by reciprocity-motivated proposers, thereby limiting surplus-destruction by proposers and promoting proposals closer to equal split. These latter are then more likely to be accepted. In Appendix A, we show that the combination of tractable reciprocity and asymmetry in conflict payoffs may predict that, while proposals in ultimatum game are aggressive and unfair, the commitments in commitment treatments are to equal splits. Thus, positive reciprocation would predict more efficiency and a more even division of the surplus in the commitment games.

Preference for equality of opportunity. Over the past decade, people have been shown to value their freedom of choice intrinsically (Falk and Kosfeld, 2006; Charness et al., 2012; Fehr et al., 2013; Bartling et al., 2014). Chlaß et al. (2019) put these values into an inequity aversion framework such that players do not only value their own (effective) options, but also care for how many they have as compared to an opponent.[21] Building on Sudgen (1998), the set of effective opportunities counts each option which yields a player distinctly different material payoff at least in some future contingency of the game and does therefore add to the player's freedom of choice. Note that a player's freedom to choose as by her effective options is a measure of the power she holds. Note also, that the above mentioned tractable reciprocity models build on such a notion, as kind acts are evaluated in terms of high-payoff opportunities which are preserved by preceding actions of the opponent. Thus, it is perhaps not surprising that models for equality of opportunity deliver predictions closely related to the tractable reciprocity models in our setting.

In the ultimatum game, there is asymmetry in the cardinalities of the sets of effective opportunities of the two players. As a second mover, the responder has at most two effective options for each proposal: : to accept or to reject. In contrast, the proposer has a range of proposals to choose from as a first mover (see Fig. A-1 in the appendix for an illustration). Thus, the responder might not only react to incapacity to reach equal outcomes induced by small proposals but also to the power asymmetry.

A player with a commitment device can voluntarily limit her own freedom of choice in the future. In addition, that same commitment device can also limit the *opponent*'s effective options. So, the responder's commitment directly regulates the proposer's options and thus the equality of effective opportunity. By refraining from committing, the responder leaves more effective opportunities to the proposer. Proposers motivated by the equality of

[21] See also (Herz and Zihlmann, 2024).

opportunity would then prefer to make more equal proposals in response to lower commitment by the responder. Where both players can commit, both the proposer's and the responder's commitments limit the proposer's effective options, and the proposer's freedom to choose may decline even more substantially. Inequality over effective opportunities states that players with greater freedom to choose, seek to compensate the opponent by granting her higher payoff; players with lower freedom of choice require such a compensation and are more willing to reject low offers.

3 Experimental design

The experimental sessions were conducted at the experimental laboratory of the Max Planck Institute of Economics in Jena, Germany. Subjects were recruited using ORSEE software (Greiner, 2015), and the performance tasks in the experiment were programmed and conducted using z-Tree software (Fischbacher, 2007).

Participants were students at Friedrich Schiller University Jena from various fields of study. There were 32 participants in each session. We organized three sessions in each treatment and, thus, the data contains observations from altogether 288 participants. When the participants arrived at the laboratory, the experimenter checked their identity and randomly assigned them a visually isolated cubicle. They received a hard-copy of the instructions, written in German, and the instructions were read out loud once the participants had read through the instructions. Participants were given an opportunity to ask questions about the protocol in private with the instructors. Thereafter, the experiment was started with a series of computerized control questions. Instructions and control questions are available in our online appendix. Once all participants had successfully answered the control questions, bargaining began. After all decisions had been submitted, subjects were paid individually according to their choices. The median earnings were 11 Euros with a minimum of 5 and a maximum of 14 euros (1 EURO = 1.29 US Dollar at the time).

As explained in Section 2, there were three alternative treatments: the *simultaneous* commitment game, the *unilateral* commitment game, and a standard *ultimatum* game. Participants were randomly assigned one of the two player roles: Player A or Player B, fixed for the entire duration of the session.

An experimental session lasted for 16 rounds; the participants played once against each participant in the opposing role in a between-subjects perfect strangers design which optimizes the number of independent observations given how many rounds have been played. Over time, observations will still become somewhat dependent due to the fact that a player learns from her own trajectory and applies this learning to the new opponent she encounters who then faces play which is inspired by past actions of other opponents, i.e. past actions by her own Player role. We could have divided each session into further matching groups but deliberately refrained from doing so: the smaller the pool of players, the more idiosyncratic and sensitive to outliers each then independent trajectory of the matching group becomes. Moreover, repeated game effects would become more of an issue in small matching groups if one wants to have several rounds of play to allow for learning. Our design opts for the largest pool of experience given the size of the laboratory while keeping observations as independent as possible so as to avoid repeated game effects, and yet also fosters convergence over time. Independent but highly idiosyncratic trajectories may diverge by design because the basis for learning is too narrow. Our repeat design still allows us to compare fully independent observations in the first round which amounts to a pure between subjects design. Once a round was played, each pair learned the payoffs for the round, commitments made (if any), the proposal and the MAO, and therefore, whether the proposal was accepted. We opted for a between-subjects design since a within-subject design would have suffered from well-known challenges related to experimenter demand effects (Zizzo, 2010, p.84).

We mostly use Nash equilibrium (and its refinements) for the purpose of providing benchmarks and of illus-

trating why one should in principle expect certain similarities in behavior between the games. We do not expect behavior to sharply comply with these predictions. Our choice of a repeated design can also be motivated with reference to these solution concepts. A central motivation for why a Nash equilibrium should arise is its nature as a steady state of a learning process where players learn about actions of others in a population and adjust their beliefs which then converge towards an equilibrium Fudenberg and Levine (2009). This motive is hardly suitable to justify the use of Nash equilibrium in the early rounds of play. More loosely, the connections in behavioral patterns between the games are more likely to arise when players have sufficient understanding of the game and each other's behavior.[22]

After the first round where observations are clearly fully independent and we have a clean between-subjects design, behavior is unlikely to have converged to a Nash equilibrium (let alone subgame perfect equilibrium) as that would require at least two degrees of mutual knowledge of self-interested rationality.

After the 16 rounds of the actual bargaining game, we used a slightly modified version of the risk aversion elicitation task (Holt and Laury, 2002) to elicit risk attitudes. Subjects also answered a battery of standardized questions allowing us to measure participants' moral judgment competence (Lind, 2008, 2021) and personal value orientation (Schwartz, 1994). The moral judgment test allows us to elicit the participants frequent or preferred ethical criteria when evaluating whether an action is ethically right.[23] It is easy to see that, arguably, social preference models in behavioral economics are associated with particular ethical criteria in the classification (see (Chlaß et al., 2019) and Table A-1). Associating these with our results contributes to an understanding whether reciprocity (Kohlberg criterion 3: intention, reciprocity, social image and social norms) or the equality of opportunity (Kohlberg criterion 5: social contract, and the equality of rights stipulated therein) are at work. We also asked participants to report gender, age, and the field of study.

4 Results

Our empirical analysis is driven by two aspects: first, the format of the dependent variable and second, the error structure. Commitments and MAOs are integer values bound within zero and twelve. Our (linear) interval regressions truncate the underlying normal distribution at 0 and 12, and truncate the values within these bounds to integers such that the underlying model generates the same range and type of values as the actual data; the interpretation remains that of a standard linear model. Payoffs in 4.2 are analyzed truncating the normal distribution at 0 and 12 for B, and 2 and 12 for A. Efficient agreement in 4.3 use simple binary logits.[24] Turning to the error structure, the data feature a panel of 16 choices for each of the 288 participants. Errors are therefore clustered at the individual level to account for repeated measurement.[25] Note that treatment effects can also be obtained by linear regressions (with biased coefficients); and a different handling of the error structure adding random, or fixed effects. Analyses of fully independent first round data use simple t-tests allowing for unequal variances.

In order to see whether a variable of interest is significantly different from zero and to know its average, we regress that variable on a constant; to know how it differs across treatments, we add treatment dummies. To see how the variable evolves dynamically, we add period, and, where it is of interest to know how treatment differences evolve, we add interaction effects of treatment and period. In the main specification of each section, we replace period by *countdown*=16-period such that treatment differences are estimated by the values they take on in the final

[22] Another key motivation for a Nash equilibrium to arise is an eductive one (Binmore, 1987). The eductive motivation might justify the use of Nash equilibrium even in the first round, but for Subgame-perfect Nash equilibrium to arise in *ultimatum*, the players should hold at least two degrees of mutual knowledge of self-interested rationality.

[23] The test is freely available for research purposes at http://moralcompetence.net; an excerpt is found in our online appendix. For a detailed description from an Economist's viewpoint, we refer to our earlier work (Chlaß et al., 2019; Chlaß et al., 2023; Chlaß and Riener, 2025). All scores are computed the exact same way as in these references which also provide results on correlations of these scores with latent variables.

[24] To support our understanding of the payoff dynamics in 4.2 , we also analyze classes of commitment types (modest: yes/no, equal: yes/no, ambitious: yes/no) using simple binary logit.

[25] Examples for experimental studies which handle the same format of the dependent variable and the same error structure (with fewer repeated observations for the same individual) in this way are (Chlaß and Moffatt, 2017; Chlaß et al., 2023).

rounds of the game where parties have had a chance to learn and behavior has had a chance to converge. Treatment differences are also significant if countdown is exchanged for period, that is, for the early rounds. The significance of descriptives in Figures 1 and 2 is assessed by one-sided tests in the direction of the descriptive pattern or result; models in our regression tables feature standard two-sided tests.

4.1 Precommitments and acceptance thresholds

Let us begin by looking at the commitments of the responders in the commitment games and their MAOs in *ultimatum*. B's MAOs and commitments aggregated over the 16 rounds are depicted in Figure 1. B-players do use precommitments to pursue higher payoffs, as predicted by the theory. Commitments (center and right panel) are set well above zero at, on average, 6.448, *p-value* = 0.000, and they are significantly above the minimal acceptable offers (MAOs) of the ultimatum game (left panel), *p-value* = 0.000. In *simultaneous* (right panel), modest commitments (below 5) by B are more frequent, *p-value* = 0.010, and ambitious commitments (greater than 6) are less frequent, *p-value* = 0.000, than in *unilateral*. The modal commitment to equal split of total surplus is more frequent in *simultaneous*, *p-value* = 0.031. Recall that our refined theory predicts that B commits to 10 in both commitment treatments. Such commitments are absent in the data, although Player A proposals with own share 10 or above are observed (39 times by 7 different participants). The observed commitments are much lower and centered around the equal split. The patterns in Figure 1 also appear in first round data, where observations are independent and no learning has yet occurred: B commitments in *unilateral* and *simultaneous* are each significantly higher than B's MAOs in *ultimatum* by Welch's one-sided t-test (when variances are unequal) with *p-values* < 0.001; B commitments in the first round are also already slightly higher in *unilateral* than in *simultaneous* with *p-value* = 0.07.

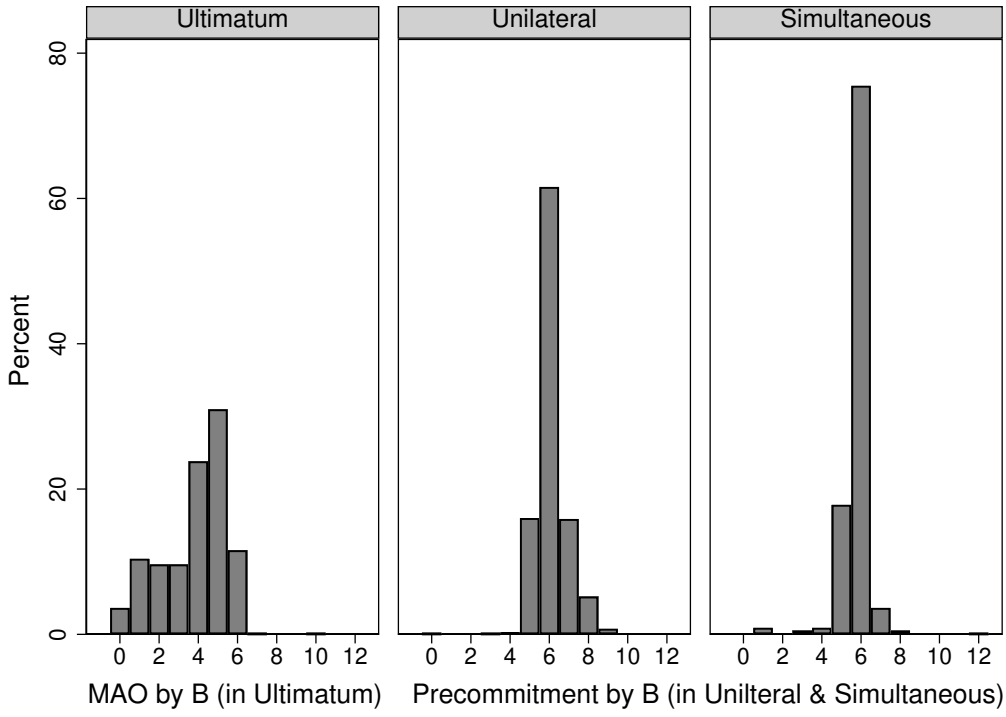


Figure 1

The aggregated patterns, illustrated in Figure 1, hide the dynamics in bargaining behavior. These dynamics, and differences in the way Player As and Player Bs exploit their pre-emptive powers, are illustrated in Figure 2.

The pre-emptive moves of the players are ordered in four classes by how large a share of pie they claim: *small* actions claim zero or one, *modest* actions claim two to four, *fair* claim five or six (where six constitutes an equal split), and *ambitious* claim seven or more. The frequency of actions in these four classes is then tracked over the 16 perfect strangers repetitions of the game. Panels on the left (right) track pre-emptive actions of Player A (B) in *ultimatum* and simultaneous (*unilateral* and *simultaneous*) commitment games, respectively. When comparing to A-players' proposals in *ultimatum* (top left panel in Figure 2), the commitments by B-players in *unilateral* are much less demanding, i.e. *p-value* = 0.000, and typically close to the equal split. Theoretically, A holds the first-mover advantage and greater bargaining power in *ultimatum*, while *B* holds this advantage in *unilateral*, where only she can commit. In the experiment, however, Player A opens the ultimatum game with a much more aggressive demand than B opens the unilateral commitment game. The difference only grows larger with experience, i.e. *p-value* = 0.030, since B does not much alter her commitment in *unilateral*, and A's proposals only get more demanding in *ultimatum*. In fact, both demands increase over time, but A's demand in *ultimatum* does so three times as much, by 0.032 *p-value* = 0.001, compared to B's commitment in *unilateral* which increases only by 0.011 per round, *p-value* = 0.012. The corresponding simple regressions which assess the significance of these empirical patterns are found in online appendix.

Player B's commitments in *simultaneous* converge towards the equal split. Modest commitments decrease over time, i.e. *p-value* = 0.014, just as ambitious commitments decrease, too, i.e. *p-value* = 0.000, whereas commitments in *unilateral* increase slightly over time as just mentioned. Commitments by B in *simultaneous* therefore remain below those in *unilateral* towards the final rounds of the game although theory predicts that there should be no difference. Commitments of As and Bs are indistinguishable in the final round of *simultaneous*, i.e. *p-value* = 0.920 although the refined theory predicts that A should not commit at all. Any just compatible outcome can be supported as an SPE in *simultaneous*, so SPE alone makes no prediction.

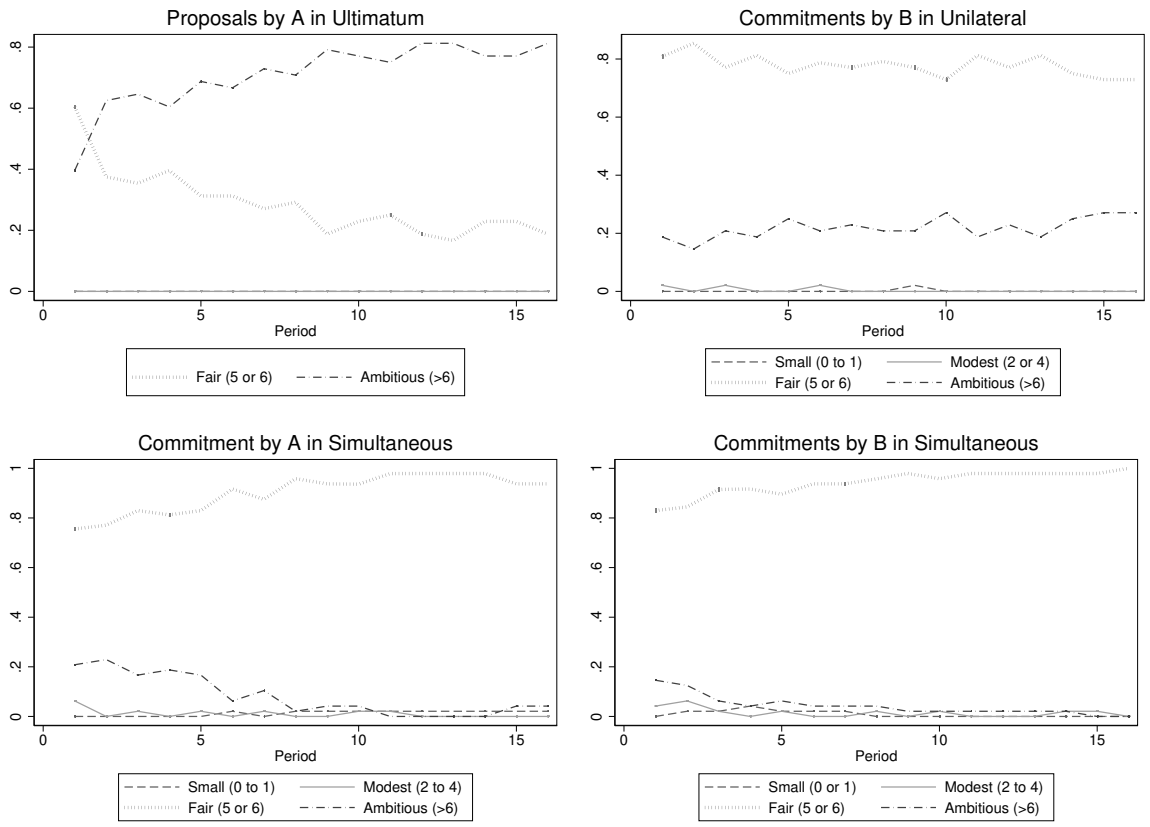


Figure 2

It turns out that also the gaps between Player Bs' commitments in *unilateral* and *simultaneous* and Bs' MAOs in *ultimatum* grow larger with experience, i.e. *p-value* = 0.003: MAOs are becoming more modest over time while the commitments in the respective treatments remain constant.

Table 1 below regresses B players' bargaining behavior – their MAOs in *ultimatum* and their commitments in *unilateral* and *simultaneous* – on treatments and the period of play. In bargaining, later periods are always of particular interest since parties have had a chance to learn about each others' behavior and have potentially converged towards equilibrium. We therefore reverse the period count such that *countdown* takes value 1 in the last period and value 16 in the first period. This way, treatment dummies directly capture the treatment differences in Player B behavior in the last period, i.e. after learning. A negative interaction of countdown with treatment indicates that Player B commitments increase over periods compared to the reference treatment. Column "All treatments" includes MAOs by Player B in *ultimatum* and commitments in *unilateral* and *simultaneous* . Commitments by B, both in *unilateral* (2.738, *p-value* = 0.000) and *simultaneous* (2.399, *p-value* = 0.000) are higher than B players' MAOs in *ultimatum*. Moreover, MAOs become lower over time since the reverse period count has a positive effect of 0.039, *p-value* = 0.004. To the contrary, commitments increase over time: *unilateral* and *simultaneous* negatively interact with countdown, i.e. -0.051, *p-value* = 0.000 and -0.048, *p-value* = 0.003, respectively. The regression in the rightmost column ("Commitment games") has data only from the two commitment games. It shows that Player B commitments are 0.340, *p-value* = 0.005, higher in *unilateral* than in *simultaneous* and that this difference persists over time since *unilateral* interacts negatively with countdown, but not significantly so, i.e. -0.003, *p-value* = 0.783.

VARIABLES	(1) All treatments B_MAO or B_commitment	(2) Commitment games B_commitment
Unilateral	2.738***	0.340***
	[0.280]	[0.121]
Simultaneous	2.399***	
	[0.267]	
countdown	0.039***	-0.009
	[0.014]	[0.008]
Unilateral*countdown	-0.051***	-0.003
	[0.014]	[0.010]
Simultaneous*countdown	-0.048***	
	[0.016]	
Constant	3.963***	6.362***
	[0.260]	[0.062]
Observations	2,304	1,536
Number of participants	144	96

Standard errors in brackets, clustered at the individual level. Interval regressions. If B had MAO or commitment of, say, 1, her MAO or commitment has lower bound 1 and upper bound 1.99. MAOs or commitments of 0 have an unknown lower bound, MAOs or commitments of 12 an unknown upper bound.

*** p<0.01, ** p<0.05, * p<0.1

Table 1: B players' MAOs & precommitments

4.2 Payoffs

Player B's strategic exploitation of precommitment is also reflected in her payoff. As predicted by theory, B's payoff is highest in the two commitment treatments. The overall difference to B's payoff in the ultimatum game is, on average, 1.24, *p-value* = 0.000. What theory does not predict is that also B's payoff in *unilateral* and *simultaneous* differ. Her overall payoff in *unilateral* is, on average, 0.323 (*p-value* = 0.012) higher than her payoff in *simultaneous*. Figure 3 shows how these differences evolve over the sixteen periods of each treatment. B's payoffs in *unilateral* and *simultaneous* – the solid lines in the center and right-hand panel of Figure 3 – lie visibly above the respective solid line in the left-hand panel from early periods on. The difference between B's payoff in *unilateral* and *simultaneous* in turn is particularly large in the first half of the rounds whereas it becomes smaller towards the end, as we note from the solid line in the center panel and the solid line in the right-hand panel. Looking at column "All treatments" in table 2, we see that after learning, in the last period, B's payoff in *unilateral* and *simultaneous* continue to lie above her payoff in *ultimatum*, namely by 2.307, *p-value* = 0.000 and 2.417, *p-value* = 0.000, respectively. Countdown interacts significantly with both treatment dummies; the difference therefore increases with learning. Looking at column "Commitment games", B's payoff in *simultaneous* in turn no longer differs significantly from her payoff in *unilateral* after learning, i.e. by 0.106, *p-value* = 0.593. Payoffs increase significantly more steeply over time in *simultaneous* than *unilateral*, the joint slope for both treatments

being 0.080, *p-value* = 0.000, with an added 0.051, *p-value*= 0.034, of steepness in *simultaneous*.

Contradicting the (refined) theoretical prediction, Player A's payoff is not lower than Player B's payoff in the commitment games. Theoretically, the gains from trade should shift from Player A to Player B when allowing commitment by B, an advantage which should prevail in the simultaneous commitment game. In fact, in the simultaneous commitment game, Player A's overall payoff is, on average, 0.351, *p-value*= 0.006, *higher* than that of B; in *unilateral*, Player A's and B's payoffs are essentially the same, *p-value* = 0.507. While B's strategic advantage over A in the commitment games therefore does not show in her payoff relative to A's, Player A's payoff in the ultimatum game features the typical marked and significant first-mover advantage.

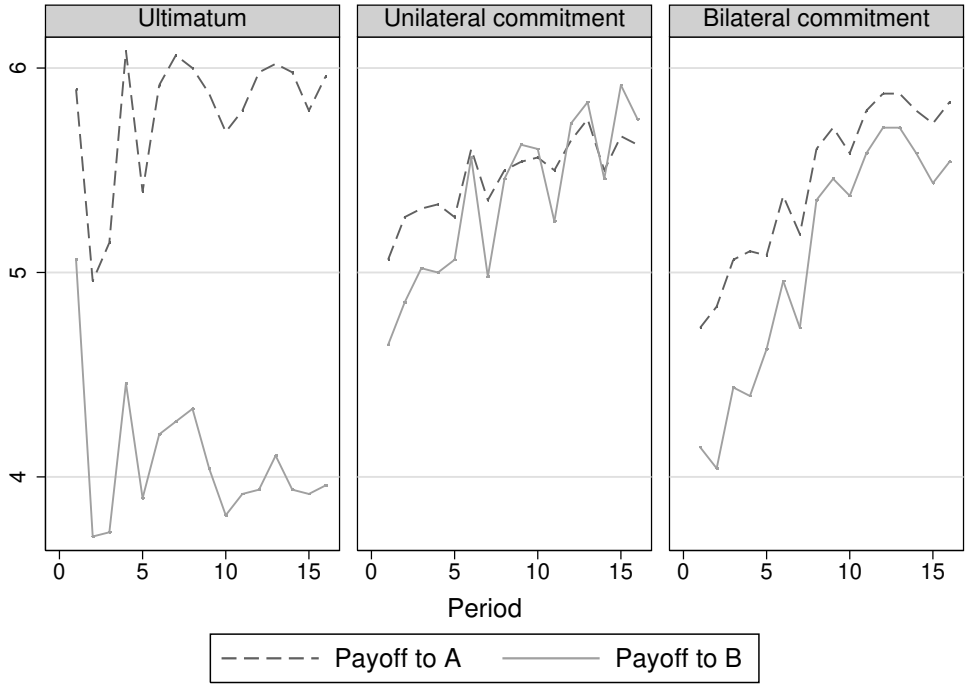


Figure 3

The commitment treatment payoffs of B are increasing (center and right panels of Figure 3), whereas her ultimatum game payoff is decreasing with experience (left panel in Figure 3). Thus, B's payoff tends to converge towards the theoretical prediction in all games, without quite reaching the predicted level in any treatment, and without really exceeding the payoff of A as theoretically predicted.

Turning to A's payoff, Figure 3 shows that it is initially higher than that of B in all treatments. In *unilateral*, A's payoff increases over time but not as steeply as B's payoff, ultimately closing the initial gap in the two parties' payoffs. The top-right panel of Figure 2 suggests that this is due to the about 20% of Player B commitments remaining steady, and Player A's partially learning to concede to those commitments, or to concede to the strategic advantage of B more generally.

In *simultaneous*, both players' payoffs rise in tandem (right panel of Figure 3) as players learn to coordinate on the fair division: almost all commitments in both player roles are eventually either 5 or 6 (proportion of commitment in class *fair* is almost 100% as illustrated by the bottom panels of Figure 2), and thus there is virtually no bargaining impasse at the precommitment stage. This improvement in coordination is reflected in increasing payoffs of both players. Table 3 shows the regressions for Player A. From column "All treatments", we see that after learning, her payoff in the commitment games is no different from her payoff in *ultimatum* since neither treatment dummy reaches significance.

VARIABLES	(1) All treatments Payoff B	(2) Commitment games Payoff B
Unilateral	2.307***	
	(0.280)	
Simultaneous	2.417***	0.106
	(0.295)	(0.198)
countdown	0.026	-0.080***
	(0.017)	(0.015)
Unilateral*countdown	-0.107***	
	(0.023)	
Simultaneous*countdown	-0.158***	-0.051**
	(0.026)	(0.024)
Constant	3.632***	5.940***
	(0.252)	(0.128)
Observations	2,304	1,536
Number of participants	144	96

Standard errors in brackets, clustered at the individual level. Interval regressions: probability mass of the normal distribution is shifted into the interval of payoffs, i.e. truncated at 0 and 12, such that the lowest value of the dependent variable which the normal distribution can generate is 0, and the highest 12.

*** p<0.01, ** p<0.05, * p<0.1

Table 2: B's payoff

A negative coefficient for the countdown variable reaffirms this to be the result of learning; the significant interaction between *simultaneous* and *countdown* shows learning is significantly more substantial in *simultaneous* than in *unilateral*. Turning to column "Commitment games", we see that A's payoff in *simultaneous* is 0.384, *p-value* = 0.010, higher than in *unilateral*, and while increasing in both commitment games, it increases significantly more steeply in *simultaneous*, i.e. by 0.050, *p-value* = 0.012.

4.3 Efficiency

By looking at the time trends in the sum of payoffs, we learn that efficiency is improving and approaching first-best in the commitment treatments unlike in the ultimatum treatment. Theory predicts that first-best should be reached in all treatments. The differences in the capacity of the various protocols to generate surplus is a surprising finding which calls for an explanation. We will return to this in the following section of the article.

To confirm these efficiency patterns, we run logit models with errors clustered at the individual level, with an (efficient) agreement dummy as the dependent variable, and treatment dummies, countdown, and their respective interactions as explanatory ones. Table 4 shows the average marginal effect for each explanatory variable; throughout, coefficients show the same significance level and sign. A marginal increase in the explanatory variable yields an increase in the probability of agreement by the size of the marginal effect displayed.

VARIABLES	(1) All treatments Payoff A	(2) Commitment games Payoff A
Unilateral	-0.186	
	(0.213)	
Simultaneous	0.210	0.384***
	(0.217)	(0.148)
countdown	-0.035**	-0.039***
	(0.015)	(0.013)
Unilateral*countdown	-0.007	
	(0.020)	
Simultaneous*countdown	-0.059***	-0.050**
	(0.022)	(0.020)
Constant	5.913***	5.732***
	(0.186)	(0.103)
Observations	2,304	1,536
Number of As	144	96

Standard errors in brackets, clustered at the individual level. Interval regressions: probability mass of the normal distribution is shifted into the interval of payoffs, i.e. truncated at 2 and 12, such that the lowest value of the dependent variable which the normal distribution can generate is 2, and the highest 12.

*** p<0.01, ** p<0.05, * p<0.1

Table 3: A's payoff

After learning, agreement in *unilateral* and *simultaneous* becomes significantly more likely than in *ultimatum*, by 21.2%, *p-value*= 0.000 and 25.7%, *value*= 0.000, respectively. In fact, this holds already at the outset. Both *unilateral* and *simultaneous* become significantly more efficient over time, by an average of 1.3%, *p-value*= 0.000 and 2.1%, *p-value*=0.000, per round, respectively. *Ultimatum*, on the other hand, does not improve in efficiency since countdown itself is not significant. We observe this significant difference in efficiency between the commitment games and *ultimatum* despite the fact that the commitment incentives of B should be analogous to the incentives of A in ultimatum. Looking at column "Commitment games" only, we see that efficiency increases similarly in the commitment games by roughly 1.1%, *p-value* = 0.000 per round, with only a slight extra of 0.7%, *p-value* = 0.087 per round in *simultaneous*.

5 Discussion

The results of the previous section depart in several ways from the hypotheses spelled out in Section 2.

In this section, we discuss the predictions of the behavioral theories (see end of Section 2) to organize the observed patters. We begin with models of inequity aversion and fairness. In the ultimatum game, we observe the usual patterns: responders reject small offers, proposers offer on average 40% of the pie (standard deviation is about 10%). Theories of inequity aversion and concern for fairness, among others, have been proposed to organize these patterns (Bolton, 1991; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000).

VARIABLES	(1) All treatments Agreement	(2) Commitment games Agreement
Unilateral	0.212***	
	(0.049)	
Simultaneous	0.257***	0.038
	(0.053)	(0.051)
countdown	-0.001	-0.011***
	(0.002)	(0.003)
Unilateral*countdown	-0.013***	
	(0.004)	
Simultaneous*countdown	-0.021***	-0.007*
	(0.004)	(0.004)
Observations	2,304	1,536
Number of Bs	144	96

Standard errors in brackets, clustered at the individual level. Binary Logit models (with a constant); marginal effect of each variable, averaged over all individuals. Throughout, coefficients of the Logits models show the same significance level and sign as their marginal effects.

*** p<0.01, ** p<0.05, * p<0.1

Table 4: Agreement rate

In other words, there is pull towards equity and fairness ideals. This does not explain, however, why the commitments of responders in *unilateral* are much more tightly concentrated around the even split than the proposals in the ultimatum game, when the incentives of the two players are almost perfectly reversed between the two games. Thus, inequity aversion or fairness concerns (alone) do not organize the patterns in our data.

Quantal-response equilibrium (McKelvey and Palfrey, 1998) is another potential classical explanation, which is known to be able to organize ultimatum game behavior (Yi, 2005). QRE also captures the observed unilateral commitments by responders. The one distinct pattern which the quantal-response equilibrium is unable to explain is the unequivocal choice of equal split commitments, by both players, in the simultaneous commitment game. Recall that, in the simultaneous move commitment game, QRE predicts that commitments by Player B stochastically dominate commitments of Player A, due to the fact that there are strict better responses to picking a perfectly matching commitment to that of B: any strategy committing more modestly or not at all and rather matching the ensuing proposal with B's realized commitment is a better-response. Therefore, extensive form QRE (alone) does not organize the patterns in our data.

A third theory which was proposed in Section 2 to contribute to an explanation is concern for equality of opportunities, i.e. concern for equal powers or equal freedom of choice (Chlaß et al., 2019).[26] According to these theories, power asymmetry both legitimizes and triggers responses to procedural unfairness: randomly assigned, unearned power may not be perceived as legitimate, and players' preferences for equality of opportunity result in calls for compensating for the inequality, making the preference for equal split and willingness to counter unequal proposals stronger. Chlaß et al. (2019) observe that such tendencies are common for those experimental

[26] See also Herz and Zihlmann (2024).

participants who often use moral arguments relating to equity principles and that is reflected in their economic choices in pie-division games.

In our commitment games, the ensuing opportunities are regulated by the commitments. In the ultimatum game, there is an asymmetry in the set of effective opportunities, but in unilateral and simultaneous commitment games, the cardinality of the proposer depends on the commitment of the responder. By refraining from making aggressive commitments but neverthess restricting opportunities for aggressive proposals by the proposer, the responder makes the opportunities more equal, preserves considerable opportunities for the proposer, and thus motivates the proposer not to make selfish proposals and herself to accept proposals as a responder. In Appendix B, we develop these arguments further and in Appendix C, we relate the observed behavior to participants moral argumentation. We show that the reasoning style associated with equality of opportunity concerns is indeed associated with the predicted behavioral patterns.

Somewhat relatedly, in tractable models of reciprocity (Cox et al., 2007, 2008), reciprocity concerns are explicitly a function of the remaining set of opportunities at a particular action node, such that if the opponent's preceding action(s) limit the highest achievable payoff in the set of opportunities, then the altruism towards the opponent will be reduced. Altruism may even reverse to spite if there are opportunities to reach equal payoffs. Given the importance of the equity reference and the role of (maximal) opportunities, it is easy to see to convince oneself of the model's explanatory power in our context. Indeed, in Appendix A we show that a tractable model of reciprocity (Cox et al., 2007, 2008) is consistent with the observed patterns.

Both classes of models predict more efficiency and more equal division of the surplus in the commitment games than in the ultimatum game. These patterns are exactly what we observe in later rounds of the interaction when participants have had ample chances to learn and behavior has converged. In the simultaneous commitment game, if both commit to exactly half of the pie, then both have exactly two effective options in the ensuing game: to accept or to reject the even split. Since effective opportunities are equal, both have every reason to accept. Still, self-interest generates the asymmetry in committing incentives: the proposer should rather wait and see. Yet, since in the later rounds, there is little strategic uncertainty about the responder's commitment, the equal split by equal commitments becomes an equilibrium. We derive the predictions of the tractable reciprocity model in Appendix A and verify its close match with the observed data. Appendix B explores concerns for the equality of opportunity in more detail. We also report the results of an additional empirical exercise where we elicit players' preferences over a rich taxonomy of ethical criteria with a long tradition in moral psychology (Piaget, 1948; Kohlberg, 1984; Lind, 2021) to see whether intention-, image-, and social norm based criteria are at work pointing toward reciprocity, or social contract reasoning in terms of the equality of rights. The latter would point toward the equality of opportunity (Chlaß et al., 2019) as it builds on an idea that moral judgment related to the fairness of the sets of opportunities matters for behavior.

Note that since reciprocity is typically about interpretation of intentions, the classification of Kohlberg would classify these models under intention-, image- and normed-based moral criteria (*Kohlberg 3*). However, the tractable model of reciprocity measures kindness with respect to changes in the set of opportunities and models reciprocity as reactions to whether high-payoff or equal-payoff opportunities are available. In that sense, reasoning at level of social contract and equality of rights (*Kohlberg 5*), in addition to *Kohlberg 3*, may be also relevant to the reciprocity model Cox et al. (2007) and especially Cox et al. (2008) which, as a revealed preference model, is agnostic about the mechanism.

There is a close analogy between our unilateral precommitent game, where the responder limits the proposer's choice set, and the principal-agent game of Falk and Kosfeld (2006) where the principal limits the agent's choice set. Falk and Kosfeld (2006) discovered behavioral patterns inconsistent with some (complete information) theories of reciprocity (Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006) and argued that the evidence

constituted novel evidence for hidden costs of control. Ellingsen and Johannesson (2008) presented a theory of type-dependent image concerns with asymmetric information about opponent type which allowed to explain these patterns.[27] One could conceivable apply this model also to the present context and account for the patterns, especially if players want to appear fair-minded (Andreoni and Bernheim, 2009).[28] It is easy to see that the almost unequivocally observed 50-50 commitments by experienced players in our commitment treatments are also highly consistent with such motives. However, such a model becomes quite complex in our three-stage games with partially simultaneous moves and, since even simpler model of reciprocity accounts for our findings, we leave the application of signaling and image models for future work.

6 Conclusion

This study investigates irrevocable and sure-to-succeed precommitments in a sequential bargaining experiment. The baseline bargaining game is the ultimatum game. In a second treatment, the responder alone can commit to automatically decline any sharing where the proposed share is below one's commitment. In a third treatment, both the responder and the proposer can make such commitments.

We observe that responders successfully use commitments to increase their payoffs. The responder payoff is higher in the commitment games than in the ultimatum game. Likewise, the payoff of the responder is higher than that of the proposer, when only the responder can commit. These comparative statics are predicted by theory.

What theory does yet also predict, and what we do *not* observe, is a perfect reversal of the first mover advantage through the introduction of responder commitment. Responders do not fully exploit the strategic advantage which their commitment entails; their commitments are not as aggressive as the proposals in the ultimatum game. When both responders and proposers can commit, behavior converges to an equal split by just compatible commitments. Although theory predicts no differences, efficiency is significantly higher in the commitment games than in the ultimatum game. There is no significant difference in efficiency between the two commitment games.

We discuss some behavioral theories and their ability to predict these findings. A tractable model of reciprocity, where players reciprocate opponent's restrictions of their opportunities, and a model of concerns for equal opportunities, where players seek and grant compensation to inequality of opportunities, are consistent with the patterns we observe. We derive the predictions and discuss the evidence in the appendix.

[27] See also Sliwka (2007). Von Siemens (2013) later showed that a reciprocity theory, to which incomplete information about types is introduced, can also explain the patterns.

[28] Ellingsen and Johannesson (2008, p.1002, fn. 23) state that willingness to appear fair-minded in the eyes of fair-minded principals could substitute willingness to appear altruistic in their model and similar results would yield.

References

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

Andreoni, J. and B. D. Bernheim (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica 77*(5), 1607–1636.

Asheim, G. B. and A. Perea (2005). Sequential and quasi-perfect rationalizability in extensive games. *Games and Economic Behavior 53*(1), 15–42.

Bade, S., G. Haeringer, and L. Renou (2009). Bilateral commitment. *Journal of Economic Theory 144*(4), 1817–1831.

Bartling, B., E. Fehr, and H. Herz (2014). The intrinsic value of decision rights. *Econometrica 82*(6), 2005–2039.

Binmore, K. (1987). Modeling rational players: Part i. *Economics & Philosophy 3*(2), 179–214.

Binmore, K., C. Proulx, L. Samuelson, and J. Swierzbinski (1998). Hard bargains and lost opportunities. *The Economic Journal 108*(450), 1279–1298.

Bolton, G. E. (1991). A comparative model of bargaining: Theory and evidence. *The American Economic Review*, 1096–1136.

Bolton, G. E. and A. Ockenfels (2000). Erc: A theory of equity, reciprocity, and competition. *American economic review 91*(1), 166–193.

Britz, V. (2013). Optimal value commitment in bilateral bargaining. *Games and Economic Behavior 77*(1), 345–351.

Charness, G., R. Cobo-Reyes, N. Jiménez, J. A. Lacomba, and F. Lagos (2012). The hidden advantage of delegation: Pareto improvements in a gift exchange game. *American Economic Review 102*(5), 2358–2379.

Chen, Z., R. Wang, and J. Zong (2024). Pre-commitment in bargaining with endogenous credibility. *Journal of Economic Behavior & Organization 227*, 106714.

Chlaß N., L. Gangadharan, and K. Jones (2023). Charitable giving and intermediation: A principal-agent problem with hidden prices. *Oxford Economic Papers 75*(4), 941–961.

Chlaß, N., W. Güth, and T. Miettinen (2019). Purely procedural preferences-beyond procedural equity and reciprocity. *European Journal of Political Economy 59*, 108–128.

Chlaß N. and P. Moffatt (2017). Giving in dictator games – experimenter demand effect or preference over the rules of the game? *CBESS Discussion Paper # 17-05*.

Chlaß N. and G. Riener (2025). Preferences for fair competition/Lying, spying, sabotaging – procedures and consequences. *Jena Economic Research Paper # 2015-16*.

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

Cox, J. C., D. Friedman, and S. Gjerstad (2007). A tractable model of reciprocity and fairness. *Games and Economic Behavior 59*(1), 17–45.

Cox, J. C., D. Friedman, and V. Sadiraj (2008). Revealed altruism 1. *Econometrica 76*(1), 31–69.

Crawford, V. P. (1982). A theory of disagreement in bargaining. *Econometrica 50*(3), 607–637.

Dufwenberg, M. and G. Kirchsteiger (2004). A theory of sequential reciprocity. *Games and economic behavior 47*(2), 268–298.

Dufwenberg, M., M. Servátka, and R. Vadovič (2017). Honesty and informal agreements. *Games and Economic Behavior 102*, 269–285.

Ellingsen, T. and M. Johannesson (2008). Pride and prejudice: The human side of incentive theory. *American economic review 98*(3), 990–1008.

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

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

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

Falk, A. and U. Fischbacher (2006). A theory of reciprocity. *Games and economic behavior 54*(2), 293–315.

Falk, A. and M. Kosfeld (2006). The hidden costs of control. *American Economic Review 96*(5), 1611–1630.

Fanning, J. and A. Kloosterman (2022). An experimental test of the coase conjecture: Fairness in dynamic bargaining. *The RAND Journal of Economics 53*(1), 138–165.

Fehr, E., H. Herz, and T. Wilkening (2013). The lure of authority: Motivation and incentive effects of power. *American Economic Review 103*(4), 1325–1359.

Fehr, E. and K. M. Schmidt (1999). A theory of fairness, competition, and cooperation. *The quarterly journal of economics 114*(3), 817–868.

Fellner, G. and W. Güth (2003). What limits escalation?—varying threat power in an ultimatum experiment. *Economics Letters 80*(1), 53–60.

Fershtman, C. and U. Gneezy (2001). Strategic delegation: An experiment. *RAND Journal of Economics*, 352–368.

Fershtman, C. and K. L. Judd (1987). Equilibrium incentives in oligopoly. *The American Economic Review*, 927–940.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics 10*(2), 171–178.

Fonseca, M. A., W. Müller, and H.-T. Normann (2006). Endogenous timing in duopoly: experimental evidence. *International Journal of Game Theory 34*, 443–456.

Fudenberg, D. and D. K. Levine (2009). Learning and equilibrium. *Annu. Rev. Econ. 1*(1), 385–420.

Gale, J., K. G. Binmore, and L. Samuelson (1995). Learning to be imperfect: The ultimatum game. *Games and economic behavior 8*(1), 56–90.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association 1*(1), 114–125.

Güth, W. and S. Huck (1997). From ultimatum bargaining to dictatorship - an experimental study of four games varying in veto power. *Metroeconomica 48*(3), 262–299.

Güth, W. and M. G. Kocher (2014). More than thirty years of ultimatum bargaining experiments: Motives, variations, and a survey of the recent literature. *Journal of Economic Behavior & Organization 108*, 396–409.

Güth, W., K. Ritzberger, and E. Van Damme (2004). On the nash bargaining solution with noise. *European Economic Review 48*(3), 697–713.

Heggedal, T.-R., L. Helland, and M. V. Knutsen (2022). The power of outside options in the presence of obstinate types. *Games and Economic Behavior 136*, 454–468.

Herz, H. and C. Zihlmann (2024). Perceived legitimacy and motivation effects of authority.

Holt, C. A. and S. K. Laury (2002). Risk aversion and incentive effects. *American economic review 92*(5), 1644–1655.

Huck, S., W. Müller, and H.-T. Normann (2002). To commit or not to commit: endogenous timing in experimental duopoly markets. *Games and Economic Behavior 38*(2), 240–264.

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

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

Kohlberg, L. (1984). *The Psychology of Moral Development*. San Francisco: Harper & Row.

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

Lind, G. (1977/2021). The moral competence test.

Lind, G. (2008). The meaning and measurement of moral judgement competence revisited-a dual-aspect model, fasko, d., w. willis (eds.), contemporary philosophical and psychological perspectives on moral development and education. cresskill.

McKelvey, R. D. and T. R. Palfrey (1998). Quantal response equilibria for extensive form games. *Experimental economics 1*, 9–41.

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

Müller, W. (2006). Allowing for two production periods in the cournot duopoly: experimental evidence. *Journal of Economic Behavior & Organization 60*(1), 100–111.

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

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

Myerson, R. B. (1991). *Game Theory: Streategy of Conflict*. Harvard university press.

Nash, J. (1953). Two-person cooperative games. *Econometrica: Journal of the Econometric Society*, 128–140.

Pattanaik, P. K. and Y. Xu (2000). On ranking opportunity sets in economic environments. *Journal of Economic Theory 93*(1), 48–71.

Piaget, J. (1948). *The Moral Judgment of the Child*. Glencoe, Illinois: Free Press.

Renou, L. (2009). Commitment games. *Games and Economic Behavior 66*(1), 488–505.

Rodriguez-Lara, I. (2016). Equity and bargaining power in ultimatum games. *Journal of Economic Behavior & Organization 130*, 144–165.

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

Schelling, T. C. (1960). *The Strategy of Conflict*. Harvard University Press.

Schwartz, S. H. (1994). Are there universal aspects in the structure and contents of human values? *Journal of social issues 50*(4), 19–45.

Selten, R. (1975). Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory 4*, 25–55.

Sliwka, D. (2007). Trust as a signal of a social norm and the hidden costs of incentive schemes. *American Economic Review 97*(3), 999–1012.

Sudgen, R. (1998). The metric of opportunity. *Economics & Philosophy 14*(2), 307–337.

Swope, K. J., J. Cadigan, and P. Schmitt (2014). That's my final offer! bargaining behavior with costly delay and credible commitment. *Journal of Behavioral and Experimental Economics 49*, 44–53.

Thaler, R. H. (1988). Anomalies: The ultimatum game. *Journal of economic perspectives 2*(4), 195–206.

Von Siemens, F. A. (2013). Intention-based reciprocity and the hidden costs of control. *Journal of Economic Behavior & Organization 92*, 55–65.

Yi, K.-O. (2005). Quantal-response equilibrium models of the ultimatum bargaining game. *Games and Economic Behavior 51*(2), 324–348.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics 13*, 75–98.

Appendix A, Reciprocity theory (Cox et al., 2007) predictions

In this section, we illustrate that a reciprocity model is consistent with our findings. We apply a simple version of the tractable model of reciprocity by Cox et al. (2007).

The utility function of player i is

$$u_i(x_i, x_j; \theta_i) = x_i + \theta(r_i)x_j$$

where

$$\theta(r_i) = \begin{cases} \theta_i, & \text{if } r_i > 0 \\ 0, & \text{if } r_i = 0 \\ -\theta_i, & \text{if } r_i < 0 \end{cases}$$

and θ_i is a positive scalar between 0 and 1. Thus, whether Player i is altruistic or spiteful towards Player $j \neq i$ depends on the reciprocity factor r_i which is defined as

$$r_i = \frac{m_i - 6}{12}$$

where m_i is maximum payoff Player i can guarantee himself given Player j's preceding action(s). That is, m_i is the maximal payoff within the remaining set of opportunities of i. Thus, tractable models of reciprocity are related to models of concerns for equality of opportunity in that both express concerns for the sets of opportunities which, in our games, may be endogenous. Remark also that 6 is the equal split reference payoff and 12 is the difference between the maximal payoff and the minimal payoff of each player in any of our games at the outset. Let us now apply this simple model to our three games to derive predictions using SPE as the equilibrium concept.

Ultimatum. Consider a proposal $(x_A, x_B) = (12 - x_B, x_B)$ by Player A. If $x_B < 6$, then a reciprocity-motivated B accepts the offer iff $x_B \geq \frac{10\theta}{1+\theta}$. It is easy to verify that this latter is the optimal proposal by Player A. For example, if $\theta = 2/3$, then the optimal proposal is 4 and this is accepted by B. More generally, no $\theta \in (0,1)$ would reject proposal $x_B = 5$ and thus it is always optimal for any Player A type to make an ambitious offer $x_B < 6$ and the equal split, $x_B = 6$, belongs to the optimal offers of type $\theta = 1$ only. Consistent with these predictions, the top-left panel of Figure 2 illustrates that 80% of the proposals by experienced proposers are ambitious, and left column of Table 1 shows that the average MAO of the experienced responders in *ultimatum* is about 4.

Unilateral. The commitment of B restricts the set of opportunities of A. If c_B is smaller (greater) than the equal split reference, then A will positively (negatively) reciprocate B. If $c_B = 6$, then A's attitude is neutral and she minds her material payoff only. Similarly to the analysis of *ultimatum*, positive and neutral cases will result in an optimal proposal equal to c_B. The case of negative reciprocation is more interesting. Assume that $c_B > 6$, then A will prefer inducing impasse payoffs over a division $(12 - c_B, c_B)$ iff $c_B > 10/(1+\theta)$. For example, if $\theta = 2/3$, any commitment strictly above the equal split, $c_B > 6$, will lead to impasse. Committing to $c_B = 6$ and inducing neutral attitudes by A towards B is the optimal commitment for B in that case. (In contrast, Player A type $\theta = 0$ will of course act in line with the rational model and not induce impasse if $c_B \leq 10$.) Consistent with these predictions, we find that 80% of the commitments of experienced Player Bs are fair (top-right panel of Figure 2).

The key contrast between these predictions is that, for a given θ, the agreements in *unilateral* are closer to equal splits than those in *ultimatum.* Any A with $\theta \geq 2/3$ will block every feasible agreement if B commits to strictly more than 6 in *unilateral*, whereas every type of B will accept a proposal of 5 in *ultimatum*. Thus, reciprocity theory predicts that aggressive precommitments will not be exploited by B to the same extent that A exploits his proposer power in *ultimatum*. A representative agent model with preference parameter $\theta = 2/3$ captures the data quite well.

Simultaneous. Let us verify that commitments to equal split, resulting in material payoff of 6, constitute equilibrium commitments. An opponent committing to an equal split acts neutrally, thus, resulting in $\theta(r) = 0$. Deviating to a higher commitment results in conflict payoff which is smaller than the payoff of 6. If Player B commits to $c_B < 6$, Player A will propose $(12 - c_B, c_B)$ resulting in a lower payoff than 6 whether B accepts of rejects the proposal. Any Player A strategy which commits to $c_A < 6$ and then proposes $(12 - c_B, c_B)$, where $c_B = 6$ will result in the same payoff as the original strategy. Thus, we have established that $c_A = 6 = c_B$ constitutes equilibrium commitment behavior. This prediction is independent of the reciprocity parameter values. The bottom panels of Figure 2 are consistent with this equilibrium prediction.

Appendix B, Equality of effective opportunities

We define an effective opportunity as an option which is *diverse* – which allows the player to generate a payout distinct from all other options.[29] To begin with, we illustrate Players' effective opportunities. In the ultimatum game, Player A makes her proposal without knowing B's MAO, and B states her MAO without knowing A's actual proposal. By stating a MAO, she can yet condition her accept-reject decision on the proposal (strategy method). Fig. A1 shows that A has 12 choice alternatives which yield her distinct monetary payoffs in some contingency of the game whereas B has *at most* two such alternatives (for each proposal).[30] This is because each MAO greater than a given proposal yields a payoff of zero, and each MAO smaller than or equal to a given proposal yields the same nonzero payoff. Therefore, the proposer has a stark advantage in effective opportunities over the responder.

In the commitment games, the capacity to reduce freedom of choice does not introduce any new distinct payoffs for either player. Thus, while an opportunity to commit, by construction, generates more opportunities, it never generates more *effective* opportunities. To the contrary, a player with an opportunity to commit can only reduce the number of effective opportunities. If B commits to reject offers below one in the unilateral commitment game, the proposer has exactly one more proposal which yields her the conflict payoff of two. This payoff is something the proposer can already ensure by offering 10. Therefore, B has reduced Player A's effective opportunities by one compared to *ultimatum*.[31] B does, of course, also reduce her own freedom to choose. But note that through commitment (to ten), she can in fact reduce Player A's effective opportunities to two – the same number she has herself, or – even more impressively perhaps – take all freedom of choice away from both players (by committing to 12).

Turning to the simultaneous commitment game, Player A's situation continues to worsen. Both players' commitments limit Player A's effective opportunities. Fig. A-2 plots Player A's commitments against Player B's and counts in each cell, the number of effective proposals available to Player A in Stage 2. We see that the proposer has no freedom of choice whatsoever left for any of the commitment combinations in the lower triangular matrix.

How do A and B react to the distribution of effective opportunities if they care for equal effective opportunities? In the ultimatum game, B should seek compensation for her lesser freedom of choice. Thus, the responder should be more willing to reject a given disadvantageous proposal than if opportunities were more equal. A should be willing to grant such a compensation.

Since B's precommitment can reduce the size of A's set of effective opportunities to that of B's, access to commitment should increase B's willingness to accept unequal offers. A should still be willing to compensate B for B's lesser opportunities. However, the higher B's commitment, and the more equal she therefore makes the number of effective opportunities, the smaller should be the compensation which A is willing to grant.

When the two players form commitments simultaneously, every more than compatible pair of commitments keeps Player A priviledged. However, every pair of just compatible commitments equalizes the number of effective opportunities: A can still take what she has committed to in Stage 2, or demand more (in which case she always obtains the conflict payoff of 2), each leading to one distinct payoff. B in turn can make a MAO equal to her

[29] The set of effective options is obtained starting from a reduced set which only contains a player's least preferred option, and then recursively adding the second-least preferred, third least preferred option and so on, until all options have been included Sudgen (1998). Clearly, we do not know a player's complete set of preference; for the analysis we therefore assume a narrowly self-interest player. Remark that we thus apply a self-interest preference-based concept of freedom of choice Pattanaik and Xu (2000). The set of effective opportunities will then be equal to the number of options which generate a distinct material payoff (Chlaß et al., 2019). Notice also that Cox et al. (2007, 2008) use this same notion when modeling reciprocal reactions to limitations of opportunities, and maximal material payoff opportunities in particular, by the opponent.

[30] Moreover, if the proposer proposes nothing, or everything, B has no freedom of choice at all, since all MAOs yield her the exact same payoff.

[31] To appreciate the connection to Cox et al. (2007, 2008), B has also eliminated the opportunity for A to reach her maximal payoff of 12. Thus, according to the the reciprocity model, A's altruism toward B should decline as a consequence. If B commits to reject offers below seven, he has reduced the opportunity for A to reach even the equal-split, thus, according to Cox et al. (2007), resulting in negative reciprocity or spite towards B.

commitment, or increase her MAO, also leading to two distinct payoffs. For the simultaneous commitment game therefore, A should still grant a compensation to B for her overall lesser decision rights, but least so among all treatments.

Appendix C, Empirical analysis of moral judgment and behavior

We regress Player B's MAOs in *ultimatum*, and her precommitments in *unilateral* and *simultaneous* on an instrument for equal effective opportunities (Chlaß et al., 2019).[32] From the moral judgment test handed out after the experiment, we elicit six moral preferences with a long-standing tradition in the field of developmental psychology. They express by which degree a subject uses a given ethical criterion to make a moral judgment, that is, to infer what is a fair course of action. According to this tradition, the following classes of criteria may contribute to an individual's judgment about the fairness of an action: if it is enforced by materialistic punishment, or by materialistic reward (*Kohlberg classes 1 & 2*); if it fosters the individual's social image, shows a good intention, respects a social norm and therefore earns approval from one's peers (*Kohlberg class 3*), or if it stems from more general societal expectations and rules, such as law and order (*Kohlberg class 4*). The individual may also use criteria such as the equality of rights stemming from the social contract (*Kohlberg class 5*); or employ universal principles such as Kant's categorical imperative (*Kohlberg class 6*). Preferences purely for the equality of effective opportunities were linked to social contract reasoning in (Chlaß et al., 2019; Chlaß and Riener, 2025), i.e. *Kohlberg class 5*. All scores, including the instrument, are computed the exact same way as in these earlier studies; scores do not differ across treatments. We control for critical demographics and, for the sake of completeness, also for subjects' personal value orientations.

Table 5 illustrates that various classes of moral reasoning explain bargaining behavior in our experiment. In the ultimatum game, for example, higher MAOs by the responder are positively associated with materialistic punishment and reward motives (*Kohlberg 3*), and social contract / equality of opportunity motives (*Kohlberg 5*), and negatively associated with concerns for the rule of law (*Kohlberg 4*). The shares proposed to oneself, then again, are negatively associated with Kantian motives (*Kohlberg 6*) and with image and social norm concerns (*Kohlberg 3*) but negatively associated with social contract / equality of opportunity motives (*Kohlberg 5*). However, experience tends to reverse the latter two effects (period x *Kohlberg 3*, period x *Kohlberg 5*). In summary, relating moral judgment to bargaining behavior confirms the relevance of many behavioral models that have been proposed to explain bargaining behavior.

We now zoom into concerns for distribution of opportunities which are relevant to the model for equality of opportunity and and which are reflected in *Kohlberg 5*.[33] From column "A's demand in UG" in Table 5, we see that *Kohlberg 5* – social contract orientation and concern for the equality of rights – is active for both players in a manner inconducive to agreement: both parties require compensation for their position of rights. Player A's demand increases in *Kohlberg 5*, i.e. 0.420, *p-value*=0.021, as does B's MAO, i.e. 0.656, *p-value* = 0.002. This pattern on Player A's side might seem inconsistent with a concern for the equality of opportunity – note, however, that A may well start out the game with the view that it is B who, by setting the MAO, defines by how much A

[32] In that first study, we could not distinguish whether the instrument measured preferences for equal information, or equal effective opportunities. Chlaß and Riener (2025) study these separately; only the preference for equal effective opportunities persists and links to the instrument. In particular, *Kohlberg class 5* is distributed the same way across studies, irrespectively of experiment and behavior; we therefore have no indication that it is not exogeneous. There are two critical controls: field of study: Law, and gender with which subjects' *Kohlberg 5* scores correlate, see the appendix to (Chlaß et al., 2019).

[33] Since reciprocity is typically about interpretation of intentions, the classification of Kohlberg would classify these models under class *Kohlberg 3*. However, the tractable model of reciprocity measures kindness with respect to changes in the set of opportunities and models reciprocity as reactions to whether high-payoff or equal-payoff opportunities are available. In that sense *Kohlberg 5* may also be relevant to the reciprocity models Cox et al. (2007) and especially Cox et al. (2008) which is agnostic to the mechanism as a revealed preference model.

can prefer one proposal over another. A Players engaging in moral reasoning of *Kohlberg 5* do, however, learn with experience to compensate B since *Kohlberg 5* $\times$ *Period* has -0.028, *p-value* = 0.043. The compound effect of *Kohlberg 5*, i.e. $0.420 - 0.028 \times$ *Period*, finally, turns negative in Period 16, and thus A's behavior eventually becomes consistent with the concern for the equality of opportunity. For Player B, the relevance of *Kohlberg 5* also weakens with experience, i.e. -0.027, *p-value* = 0.08, but less significantly than for A. We therefore conclude that the effect reverses for Player A, and remains robust for Player B. Turning to column "A's demand in unilateral", B's commitment has a negative effect on the demand of A as the rational theory predicts. Controlling for B's commitment, A's demand is indeed lower, the more she cares for *Kohlberg 5*, i.e. -0.935, *p-value* = 0.007. Social contract orientation does therefore indeed lead to a compensation of Player B who has fewer opportunities. The higher B's commitment, however, and the more she therefore reduces A's effective opportunities, the less Player A compensates. This is visible in the interaction *opponent's commitment* $\times$ *Kohlberg 5* which is positive, i.e. 0.157, and significant *p-value* = 0.015 on A's demand for herself. Looking at 'A's commitment in simultaneous", Player A starts out making lower commitments, the stronger her social contract orientation, i.e. -0.408, *p-value* = 0.035. That is, she starts out intending to compensate B. The effect reverses, however, with experience since *Kohlberg 5* interacts positively with experience, i.e. 0.028, *p-value* = 0.035. The compound effect of *Kohlberg 5* turns positive in Period 15. For Player B, ethical concerns crowd out altogether once a precommitment device reduces inequality in effective opportunities, see our online appendix. This is an interesting observation in itself. Overall, we observe a substantial amount of learning. First, as to how a Player assesses the bargaining situation by her own ethical standards; and second, how a Player reacts to the behavior of the other over rounds of repetition.

To summarize, moral judgment is associated with behavior and treatment effects in interesting and significant ways. Of the moral reasoning which we observe, *Kohlberg class 5* is of particular interest as it is this class which indicates concerns for the equality of opportunity. Indeed, *Kohlberg 5* is active in a way which could hamper the reversal of the strategic advantage from *ultimatum* to *unilateral*. One way to read Table 5 is as follows: B demands more for herself in *ultimatum* where she has a disadvantage in effective opportunities, the stronger her *Kohlberg 5* reasoning. Her *Kohlberg 5* reasoning is no longer active in the commitment games (see our online appendix) where B has obtained the power to reduce her disadvantage in effective opportunities. Player A's *Kohlberg 5* reasoning moderates the extent to which she gives in to B's commitment in *unilateral*; in *simultaneous*, her *Kohlberg 5* reasoning lessens her commitments. If A submits lower commitments, then B does so, too (see our online appendix) which implies lesser conflict.

VARIABLES	A'S DEMAND UG	B'S MAO UG	A'S DEMAND UNILATERAL	A'S COMMITMENT SIMULTANEOUS
Constant	8.979***	2.453**	8.278***	6.566***
	(0.711)	(1.190)	(0.808)	(0.480)
Kohlberg class 1	0.068	0.115	−0.013	−0.069
	(0.126)	(0.314)	(0.032)	(0.091)
Kohlberg class 2	0.221	0.868**	0.107	−0.384***
	(0.182)	(0.379)	(0.085)	(0.103)
Kohlberg class 3	−0.525***	−0.453	0.490	0.388**
	(0.139)	(0.327)	(0.567)	(0.179)
Kohlberg class 4	0.002	−1.119***	−0.209***	0.317**
	(0.132)	(0.393)	(0.072)	(0.161)
Kohlberg class 5	0.420**	0.656***	−0.935***	−0.408**
	(0.182)	(0.212)	(0.347)	(0.194)
Kohlberg class 6	−0.509***	−0.353	−0.033	0.036
	(0.131)	(0.247)	(0.045)	(0.112)
Kohlberg class 3	0.035***	0.015	0.006	−0.028*
× Period	(0.013)	(0.010)	(0.006)	(0.015)
Kohlberg class 5	−0.028**	−0.027*	0.003	0.028**
× Period	(0.014)	(0.015)	(0.005)	(0.014)
opponent's commitment			−0.260**	
			(0.124)	
opponent's commitment			−0.068	
× Kohlberg 3			(0.103)	
opponent's commitment			0.157**	
× Kohlberg 5			(0.065)	
opponent's commitment			−0.020*	
× Period			(0.010)	
Period	0.029***	−0.043***	0.110*	−0.031**
	(0.008)	(0.014)	(0.062)	(0.015)
Risk aversion	−0.211***	0.298**	−0.019*	−0.022
	(0.059)	(0.144)	(0.016)	(0.046)
Gender:female	−0.003	−0.088	0.024	−0.016
	(0.080)	(0.143)	(0.060)	(0.061)
Age	−0.005	0.017	0.001	0.002
	(0.018)	(0.019)	(0.011)	(0.017)
Universalism	0.042	0.003	−0.034	0.060
	(0.047)	(0.060)	(0.030)	(0.064)
Benevolence	−0.043	−0.041	−0.068*	−0.047
	(0.040)	(0.068)	(0.035)	(0.036)
Power	0.026	−0.007	−0.056*	−0.010
	(0.054)	(0.092)	(0.030)	(0.023)
Achievement	0.018	0.049	0.001	0.008
	(0.047)	(0.078)	(0.024)	(0.052)
Fields of study	YES	YES	YES	YES
observations	768	768	768	768

Standard errors in brackets, clustered at the individual level. Interval regressions, see Notes to Tables 1,2 & 3.

$***\ p < .01$, $**\ p < .05$, $*\ p < .10$.

Table 5: Player A's demand (proposal to himself) and B's demand (MAO) in the ultimatum game, and A's precommitments in the commitment games.

Appendix D, Number of effective opportunities: Ultimatum Game

Player A \ Player B	0	1	2	3	4	5	6	7	8	9	10	11	12
0	0 12	0 2	0 2	0 2	0 2	0 2	0 2	0 2	0 2	0 2	0 2	0 2	0 2
1	1 11	1 11	0 2	0 2	0 2	0 2	0 2	0 2	0 2	0 2	0 2	0 2	0 2
2	2 10	2 10	2 10	0 2	0 2	0 2	0 2	0 2	0 2	0 2	0 2	0 2	0 2
3	3 9	3 9	3 9	3 9	0 2	0 2	0 2	0 2	0 2	0 2	0 2	0 2	0 2
4	4 8	4 8	4 8	4 8	4 8	0 2	0 2	0 2	0 2	0 2	0 2	0 2	0 2
5	5 7	5 7	5 7	5 7	5 7	5 7	0 2	0 2	0 2	0 2	0 2	0 2	0 2
6	6 6	6 6	6 6	6 6	6 6	6 6	6 6	0 2	0 2	0 2	0 2	0 2	0 2
7	7 5	7 5	7 5	7 5	7 5	7 5	7 5	7 5	0 2	0 2	0 2	0 2	0 2
8	8 4	8 4	8 4	8 4	8 4	8 4	8 4	8 4	8 4	0 2	0 2	0 2	0 2
9	9 3	9 3	9 3	9 3	9 3	9 3	9 3	9 3	9 3	9 3	0 2	0 2	0 2
10	10 2	10 2	10 2	10 2	10 2	10 2	10 2	10 2	10 2	10 2	10 2	0 2	0 2
11	11 1	11 1	11 1	11 1	11 1	11 1	11 1	11 1	11 1	11 1	11 1	11 1	0 2
12	12 0	12 0	12 0	12 0	12 0	12 0	12 0	12 0	12 0	12 0	12 0	12 0	12 0

Notes. Player A's proposals and Player B's MAOs. **Cells contain the respective payoff for each strategy combination**. A proposal of Zero yields Player B the same payoff for all MAOs; a proposal of 12 Euros does so, too. In both cases, she has zero freedom of choice. For all other proposals, she can reach two different outcomes, and has two effective opportunities.

Figure A-1.

Appendix E, Number of effective opportunities: simultaneous commitment game

Player B's commitment in Stage 1

Player A's commitment (demand) in Stage 1	0	1	2	3	4	5	6	7	8	9	10	11	12
0	13	13 – 1	13 – 2	13 – 3	13 – 4	13 – 5	13 – 6	13 – 7	13 – 8	13 – 9	13 – 10	13 – 11	13 – 12
1	13 – 1	13 – 1 – 1	13 – 1 – 2	13 – 1 – 3	13 – 1 – 4	13 – 1 – 5	13 – 1 – 6	13 – 1 – 7	13 – 1 – 8	13 – 1 – 9	13 – 1 – 10	13 – 1 – 11	∅
2	11	10	9	8	7	6	5	4	3	2	1	∅	∅
3	10	9	8	7	6	5	4	3	2	1	∅	∅	∅
4	9	8	7	6	5	4	3	2	1	∅	∅	∅	∅
5	8	7	6	5	4	3	2	1	∅	∅	∅	∅	∅
6	7	6	5	4	3	2	1	∅	∅	∅	∅	∅	∅
7	6	5	4	3	2	1	∅	∅	∅	∅	∅	∅	∅
8	5	4	3	2	1	∅	∅	∅	∅	∅	∅	∅	∅
9	4	3	2	1	∅	∅	∅	∅	∅	∅	∅	∅	∅
10	3	2	1	∅	∅	∅	∅	∅	∅	∅	∅	∅	∅
11	2	1	∅	∅	∅	∅	∅	∅	∅	∅	∅	∅	∅
12	1	∅	∅	∅	∅	∅	∅	∅	∅	∅	∅	∅	∅

Notes. Player A's and Player B's commitments in Stage 1. **Cells count how many options (proposals) the proposer has left in Stage 2 per combination of commitment.** Each additional unit of commitment (demand) by a player takes one option from A in Stage 2. Where Player A has only a single potentially successful proposal left, we indicate this as an effective opportunity since she can still reach two payoffs: the conflict payoff of Two, and the one implied by the successful proposal. Note, too, that we count down from thirteen, not twelve: for an A commitment smaller or equal to Two, a proposal of 10 for Player B is redundant with the conflict payoff for all MAOs smaller than 10 and therefore does not constitute an effective opportunity. For A commitments beyond Two or MAOs beyond 10, the conflict payoff in Stage 2 yields a distinct payoff and therefore constitutes an effective opportunity. For the illustration above therefore, we always keep conflict in the set of options. Note finally, that where only conflict is possible in Stage 2, Player A has zero freedom to choose which we indicate by the empty set, i.e. ∅.

Figure A-2.

Appendix F, The taxonomy of moral argumentation (Kohlberg, 1984, Lind, 1977-2021, Piaget, 1948) used in Section 6 and its correspondence with a row of social preference models.

preconventional argumentation	preference models
Kohlberg 1. Orientation toward punishment and obedience, physical and material power. Rules are obeyed to avoid punishment. ***Kohlberg 2.*** Naïve hedonistic orientation. The individual conforms to obtain rewards.	(...)
conventional argumentation	**preference models**
Kohlberg 3. Orientation toward inter-individual mutual relations. "Good boy/girl" orientation to win the approval and maintain the expectations of one's immediate group. The individual conforms with norms and expectations and shows good intentions to avoid disapproval. One earns approval by being "nice".	guilt aversion (Battigalli and Dufwenberg 2007), inequity aversion (Fehr and Schmidt 1999, Bolton and Ockenfels 2000), social norms (Bicchieri, 2006; Krupka & Weber, 2013) , reciprocal preferences (Falk and Fischbacher 2006, Dufwenberg and Kirchsteiger 2007), preferences for equal expected payoffs (Bolton et al. 2005); preferences for kind procedures (Sebald 2010); image (Ellingsen & Johannesson, 2008; Andreoni & Bernheim, 2009)
Kohlberg 4. Orientation toward law and order, and in particular societal expectations and moral rules from outside one's immediate peer group since these maintain and ensure the continuity of the social order.	rule-following
postconventional argumentation	**preference models**
Kohlberg 5. Orientation toward the social contract. Duties are defined by the social contract and the equality of rights stipulated therein. Emphasis is on mutual commitment and obligation in a liberal democratic basic order.	Chlaß et al.'s (2019) purely procedural preferences: equality of decision rights and information rights;
Kohlberg 6. Orientation toward universal ethical principles of conscience such as Kant's categorical imperative. Rightness of an act is derived from abstract and consistent ethical principles such as the inalienability of human rights, the free will, and individuals' freedom to choose. Ethical principles are a priori truths inherent in rational beings as laid down by Kant's categorical imperative.	Alger and Weibull's (2013) Homo Moralis

TABLE A1: SIX CATEGORIES OF LAWRENCE KOHLBERG'S TAXONOMY OF MORAL ARGUMENTATION AND A LIST OF ECONOMIC PREFERENCE MODELS BUILT ON THE RESPECTIVE CRITERIA LISTED IN EACH CATEGORY.

Notes. Sources: Kohlberg (1984). Economic preference models listed are *preferences for kind procedures.* Sebald, A. (2010), Attribution and Reciprocity, Games and Economic Behavior, 68, pp. 339-352; *preferences for equal expected payoffs.* Bolton, G., Brandts, J.,Ockenfels A. (2005), Fair Procedures: Evidence From Games Involving Lotteries, Economic Journal, 115, pp. 1054-1076; *guilt aversion.* Battigalli, P. and M. Dufwenberg (2007), Guilt in Games, American Economic Review, Papers and Proceedings, 97, pp. 170-176; *reciprocal preferences* Falk, A., Fischbacher, U. (2006), A Theory of Reciprocity, Games and Economic Behavior, 54, pp. 293-315, Dufwenberg, M., Kirchsteiger, G. (2004), A Theory of Sequential Reciprocity, Games and Economic Behavior, 47, pp. 268-98. *inequity aversion* Fehr and Schmidt (1999), Bolton, G., Ockenfels A. (2000) ERC - A Theory of Equity, Reciprocity and Competition, American Economic Review, 90, pp. 166-193; Fehr, E., Schmidt, G. (1999), A Theory of Fairness, Competition and Cooperation, Quarterly Journal of Economics, 114, pp. 817-868; *social norms.* Bicchieri, C. (2006). The Grammar of Society, Cambridge, UK: Cambridge University Press, Krupka, E. L., and Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary?. Journal of the European Economic Association, 11(3), pp. 495-524; *image.* Ellingsen, T. and Johannesson. M. (2008), Pride and Prejudice: The Human Side of Incentive Theory, American Economic Review, 98 (3), pp. 990–1008, Andreoni, J. and Bernheim, B.D. (2009), Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects, Econometrica, 77(5), pp. 1607-1636; *purely procedural preferences*. Chlaß N., Güth, W., Miettinen, T. (2019), Purely procedural preferences – Beyond procedural equity and reciprocity, European Journal of Political Economy, 59, pp. 108-128 *Homo Moralis.* Alger, I. and Weibull, J.W. (2013), Homo Moralis - Preference Evolution Under Incomplete Information and Assortative Matching, Econometrica 81(6), pp. 2269-2302.