CESifo WORKING PAPERS

8980
2021

March 2021

Too Big to Prevail: The Paradox of Power in Coalition Formation

Changxia Ke, Florian Morath, Anthony Newell, Lionel Page

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 8980*

Too Big to Prevail: The Paradox of Power in Coalition Formation

Abstract

In standard coalition games, players try to form a coalition to secure a prize and a coalition agreement specifies how the prize is to be split among its members. However, in practical situations where coalitions are formed, the actual split of the prize often takes place after the coalition formation stage. This creates the possibility for some players to ask for a renegotiation of the initial split. We predict that, in such situations, a player can suffer from being "too strong". Our experimental results confirm that, when the actual split of the prize is delayed, a player's strength can turn into a strategic disadvantage: a greater voting power in forming a winning coalition is undermined by the threat of being overly powerful at the stage when a split is determined. This result is relevant to many real world situations where "too strong" players find it paradoxically hard to partner with weaker players to win the game.

JEL-Codes: C710, C920, D720, D740.

Keywords: Shapley Value, (non) binding agreement, balance of power, communication.

Changxia Ke
Queensland University of Technology
Brisbane / QLD / Australia
changxia.ke@qut.edu.au

Florian Morath
University of Innsbruck
Innsbruck / Austria
florian.morath@uibk.ac.at

Anthony Newell
Queensland University of Technology
Brisbane / QLD / Australia
anthony.newell@qut.edu.au

Lionel Page
University of Technology
Sydney / NSW / Australia
lionel.page@uts.edu.au

March 23, 2021
We are grateful to Ralph Bayer, Guillaume Frechette, Ben Greiner, Rebecca Morton, Henrik Orzen, seminar and conference participants at Sun Yat-sen University, Wuhan University, Darmstadt University, ZEW Mannheim, the 2018 meeting of the Verein für Socialpolitik (Freiburg), and the 2018 APESA meeting (Brisbane) for their helpful comments.

"'Bargaining power,' 'bargaining strength,' 'bargaining skill' suggest that the advantage goes to the powerful, the strong, or the skillful. ... These qualities are by no means universal advantages in bargaining situations; they often have a contrary value." Schelling (1980)

1 Introduction

This paper considers the role of asymmetries in the players' power when players form coalitions in order to allocate a resource. We show experimentally that greater bargaining power can make the player worse off when coalition agreements are in part negotiated after the formation of the coalition.

Coalitional game theory typically considers situations where players form coalition agreements which are *binding*: such agreements define how the coalition's prize[1] will be shared, without any room for renegotiation at a later time. In such situations, players with more voting power are attractive coalition partners. Their power can increase the coalition's chances of winning. However, a key aspect of real world coalitional games is that agreements are most often not strictly binding. Future renegotiation can therefore not be ruled out and the decision on the prize division is in practice finalised after the formation of the coalition itself. This fact critically changes the appeal of powerful players as coalition partners. When considering joining a strong player's coalition, weak players have now to weigh the greater chance of being in a winning coalition versus the prospect of possibly having weaker intra-coalition bargaining power in the future. We show that without the possibility to pre-commit to a prize division, an increase in the power of the strong player can, paradoxically, be disadvantageous and make her less likely to form coalitions than a weak player.

To investigate this question we design an experimental weighted majority game with one powerful ('strong') player and three standard ('weak') players. The notion of 'power' is mapped in the number of votes assigned to a player. We analyse the effect of an increase in the strong player's voting rights on the coalition formation outcomes in three different

[1] In coalitional games, players aim to get a share of a prize by forming a (winning) coalition with other players. A typical example is a situation where different parties aims to form a majority coalition in a Parliament, with the prize being the formation of the government.

scenarios: (i) a baseline situation in which the players simultaneously vote on coalitions together with a prize division (treatments with *binding agreement*); (ii) a situation in which the players first vote on preferred coalitions and afterwards vote on a prize division within the winning coalition (treatments with *delayed agreement* on prize division); and (iii) a variant of the treatments with delayed agreement in which the individuals can communicate their intentions of how to divide the prize at the coalition formation stage (treatments with *delayed agreement with communication*). The latter can be seen as an intermediate treatment where, as in many real-world applications, the players can indicate their split proposals ex ante even when bargaining about the prize split after the coalition has formed cannot be avoided.

Since it is well known in bargaining theory that the exact specification of the bargaining process can be crucial for equilibrium outcomes, we impose as little procedural rules as possible and let the experimental subjects make simultaneous proposals and vote on these proposals simultaneously until they reach an agreement. One main advantage of this approach is that we do not favour a certain coalition/player by the rules of the game; moreover, we are able to relate differences in voting outcomes to asymmetries in the players votes. For theory predictions, we rely on coalitional game theory to establish our benchmark predictions on the outcome of the coalition formation game at a given stage, where we measure the players' relative power by their Shapley values. When agreement occurs in two separate stages (in the treatments with *delayed agreement*), we assume that the players form expectations about their bargaining power (payoffs) at the prize-split stage based on their Shapley values within different winning coalitions. Anticipating their expected subgame payoffs, they propose and vote on the coalitions to be formed using backward induction.[2] The experiment sheds light on whether this intuitive logic is supported by the empirical data and whether it may result in a "paradox of power" in the coalition formation game.

[2]As an index of power, a higher Shapley value is naturally interpreted as predicting a higher average payoff and/or a higher probability to be present in the winning coalition. Previous studies have empirically tested the Shapley value as an index of power. Several studies have stressed the limitation of an abstract measure that does not take into account the full complexity of institutions and political spaces which provide the underlying structure to real world coalitional games (Felsenthal et al., 1998; Leech, 2002; Gelman et al., 2004), whereas other studies have argued that the Shapley value is a good predictor of actual outcomes such as the allocation of the EU budget between member states (Kauppi and Widgrén, 2004). Since our focus is on treatment comparisons within a given institutional framework, we use the Shapley value as a reasonable benchmark for the subgame payoffs that highlights the intuitive mechanism behind the predicted treatment effects.

Our main findings are as follows. First, with *binding agreement*, the results of the game are roughly in line with the relative power of the players as measured by their Shapley value. The higher number of votes of the strong player translates into a higher likelihood to be in the winning coalition, as the strong player can be pivotal more often in the different possible coalitions. However, as long as the strong player does not hold a majority by herself, increasing the strong player's votes does not significantly affect her chance to be part of a winning coalition with *binding agreement*. Second, in case of *delayed agreement* and in line with our theory prediction, the strong player's likelihood to be part of the winning coalition stays high if she is not "too strong" but it decreases significantly if she becomes too strong in the sense that she would have the full bargaining power in minimal winning coalitions (coalitions with one weak player). Consequently, the absence of binding agreement significantly reduces a strong player's average payoff once her voting power becomes too high. Third, communication at the coalition formation stage does not mitigate the commitment problem of the strong player in the absence of binding agreements if the strong player is too powerful. By and large, the experimental results confirm the importance of binding agreements for "too strong" players' coalition formation capacity.

The novelty of our study is to investigate, in a controlled experiment, a coalitional game as taking place over two stages, delaying the final split within the coalition to a second stage.[3] Such a two-stage approach can be related to the study of proto-coalition formations (Baron and Diermeier, 2001; Diermeier et al., 2008). Proto-coalition games reflect the dynamics of formation of many governments whereby a "formateur" is selected to propose to a subset of players to form a majority coalition. This majority coalition then negotiates a prize split with a unanimity rule. In our two stage coalitional game, the negotiation within the winning coalition does not necessarily require agreement of all members of this coalition, leading players to take into account their actual bargaining power in the coalition when deciding to

[3]From a game-theoretical perspective there is a large body of work on coalition formation that highlighted interesting conceptual questions such as the stability of coalitions and the issue of farsightedness, both using cooperative and non-cooperative approaches. For an early paper on stability concepts see Hart and Kurz (1983); recent overviews of the theory of coalition formation are Ray (2007); Bloch and Dutta (2011). Our setting differs from Owen (1972)'s composition of games, where the population is divided in subgroups (e.g. constituency) and each subgroup selects a representative (e.g. Congressman). These representatives then play a coalitional game with each other in a larger game (e.g. Congress).

join it.[4] Our experimental results provide evidence that such intra-coalition considerations may lead weak players to shy away from forming coalitions with too powerful partners.

The paradoxical effect of power we observe also echoes other documented phenomena. In the chairman paradox described by Farquharson (1969) a chairman with a tie-breaking vote may be less likely to get his preferred options when faced with sophisticated voters who ensure that the chairman is not pivotal in the choice of his preferred option. It also echoes the theoretical result from Acemoglu et al. (2008) in coalition games where coalitions can be at any time contested by sub-coalitions. In their setup, a strong player may be more likely to be removed from a majority coalition. It is also related to the notion of "balance of power" whereby countries ally together to counteract a growing power and keep the regional balance. Political scientists have recorded that, when faced in a local conflict with the options of bandwagoning (allying with the strong state) or balancing (allying with the opposite coalition), states have often opted for the latter option throughout history to avoid the emergence of a hegemonic power (Kaufman et al., 2007).

Following the traditional cooperative game theory approach, our experimental coalition formation game avoids assumptions on specific bargaining institutions. In an early experiment, Fiorina and Plott (1978) compare different solution concepts in an unstructured bargaining model with face-to-face communication.[5] A recent experiment by Tremewan and Vanberg (2016) analyses resource divisions and their stability in an unstructured bargaining game in continuous time, involving three symmetric players. More specifically, our paper relates to the few experimental studies of weighted majority games that investigate the predictive power of the Shapley value in controlled environments. These studies have also provided support to the use of the Shapley value as a measure of power; see Montero et al. (2008) for the effect of an enlargement of the voting body and Geller and Mustard (2007) for changes in the distribution of votes. Nash et al. (2012) consider coalition formation in

[4]Such strategical concerns have been explicitly suggested by Robert Aumann describing the strategy of the leader called to form a majority government in Israel: "the hypothesis that the leader aims to maximize his Shapley value seems a reasonable hypothesis to test, and it works not badly." (van Damme, 1998).

[5]Structured bargaining models are most notably analysed in models of legislative bargaining where the prize division proposals are made sequentially. This literature has focused on the impact of different voting rules and the role of proposer power (Frechette et al., 2005; Drouvelis et al., 2010; Diermeier, 2011); see also Agranov and Tergiman (2014) and Baranski and Kagel (2015) on the impact of pre-play communication on proposer power.

a sequential game in which players can transfer their bargaining power to another player and find support for resource allocations in line with the Shapley value in a repeated-game setting. Our paper, instead, highlights the commitment problem that a strong player may face, turning formal voting power into a strategic disadvantage.[6]

2 Theoretical and experimental framework

2.1 Basic setup

The experimental framework is based on a weighted majority game with four players. The set of players is denoted by N. Each player $i \in N$ is endowed with a number of votes $m_i \in \mathbb{N}_+$. One player is called the "strong" player S and holds m_S votes. The three other players are called "weak" players W, each with a number of votes $m_W < m_S$. We assume that $m_S < 3m_W$, that is, the strong player holds less than 50% of the total votes. All players' votes are commonly known. There is a fixed amount of money $b = 1$ (the prize) that the players have to divide among themselves. A split of the prize is denoted by $\boldsymbol{\beta} = (\beta_1, ..., \beta_4)$, where $\beta_i \geq 0$ represents player i's payoff under this split.

To divide the prize the players must form coalitions. A coalition C is a subset (of the set N) of players. A majority coalition is defined as a coalition $C \subseteq N$ with $\sum_{i \in C} m_i > \frac{1}{2} \sum_{i \in N} m_i$, that is, as a coalition that includes more than 50% of the total votes. The set of majority coalitions is $\mathbf{C}_M$. By winning coalition we refer to the majority coalition $C \in \mathbf{C}_M$ that is selected to divide the prize.

The main experimental variation relates to the process of how to divide the prize: we vary whether there is commitment to a certain allocation at the coalition formation stage.

[6] The importance of (commitment to) power sharing is also discussed in political science in the context of political leadership and survival (Myerson, 2008; Magaloni, 2008; Svolik, 2009). This literature emphasises the moral hazard problem of the leader (dictator) once elected and the role of elections and delegation to political organisations (parties) as means to solve the commitment problem. A related moral hazard problem emerges in our framework in the absence of binding agreement where a too powerful player, once being part of the winning coalition, cannot be prevented from exploiting her power.

Binding agreement In the first variant, coalition formation involves binding agreement on how to split the prize; the rules map a version of a "weighted majority" game.[7] Each player $i \in N$ proposes a majority coalition $C \in \mathbf{C}_M$ together with a split $\boldsymbol{\beta}$ of the prize among the coalition members $j \in C$. The proposed split must satisfy $\beta_j > 0$ if and only if $j \in C$, and $\sum_{j \in C} \beta_j = 1$. These proposals $(C, \boldsymbol{\beta})$ are made simultaneously and independently. Denoting the set of proposed coalitions that include i by $\mathbf{C}_P^i$, each player i observes all proposals $(C, \boldsymbol{\beta})$ with $C \in \mathbf{C}_P^i$ and votes for one of these proposals. If a proposal $(C, \boldsymbol{\beta})$ receives the votes of all players $j \in C$, this coalition/split is implemented, player i receives a payoff of β_i, and the game ends. (Since only coalitions $C \in \mathbf{C}_M$ can be proposed, the coalition that finds agreement of all members must have necessarily obtained a majority of votes.) Otherwise, the stage is repeated with probability 95% and the game ends with zero payoffs for all players with probability 5%.[8]

This structure maps a four-player coalitional game $(N; v)$ where the characteristic function v is given by $v(C) = 1$ if $C \in \mathbf{C}_M$ and $v(C) = 0$ otherwise. With a slight abuse of notation, the set of majority coalitions consists of the following types of coalitions:

$$\mathbf{C}_M = \{(S, W), (S, W, W), (W, W, W), (S, W, W, W)\} \tag{1}$$

where (S, W), for instance, represents a coalition between the strong and one weak player (there are, hence, three such coalitions). Intuitively, a weak player either needs the strong player or two other weak players to reach majority, whereas the strong player needs one of the three weak players only. This holds for all $m_S \in (m_W, 3m_W)$: within this range, an increase in the number of votes of the strong player (m_S) neither affects $\mathbf{C}_M$ nor the characteristic function v in the variant with binding agreement. Correspondingly, the players' Shapley values–measuring their bargaining power by how often they can be pivotal in winning coalitions–are:

$$Sh(N; v) = \left(\frac{1}{2}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right), \tag{2}$$

[7]The weighted majority game is given by $[q; m_1, ..., m_4]$ with quota $q = \left(1 + \sum_{i \in N} m_i\right)/2$.

[8]We follow here the approach of Nash et al. (2012) who used a slightly more stringent 10% discontinuing probability in their experiment.

for all (m_S, m_W) with $m_S \in (m_W, 3m_W)$.[9] This observation provides the benchmark for the effect of increasing the strong player's votes, which we contrast with an increase in m_S in a case without ex ante commitment at the coalition formation stage.

Delayed (split) agreement In the second variant of the game, the players cannot commit to a split at the point when coalitions have to be formed. The allocation of the prize takes place in a dynamic two-stage game, which separates coalition formation from the division of the prize.[10]

Stage 1 of the game is the coalition formation stage. Each player i proposes a majority coalition $C \in \mathbf{C_M}$; the proposals are made simultaneously and independently. Then, each player i observes all proposed coalitions $C \in \mathbf{C}_P^i$ that include herself and votes for one of the coalition proposals $C \in \mathbf{C}_P^i$. If a proposed coalition $\tilde{C}$ receives the votes of all players $j \in \tilde{C}$, this coalition is formed and the game proceeds to stage 2. Otherwise, stage 1 is repeated with probability 95% and the game ends with zero payoffs for all players with probability 5%.

In stage 2 of the game, the prize has to be divided within the winning coalition $\tilde{C} \in \mathbf{C}_M$ that has been formed in stage 1. Each player $i \in \tilde{C}$ makes a proposal $\boldsymbol{\beta}$ to split the prize among the coalition members $j \in \tilde{C}$ (or a subset of $\tilde{C}$); these proposals are made simultaneously and independently.[11] Each coalition member i observes all split proposals that include herself (that is, where $\beta_i > 0$) and votes for one of the proposed splits. A split $\boldsymbol{\beta}$ is implemented if all players k included in the proposal (with $\beta_k > 0$) vote for it and, together, this subcoalition holds a "supermajority" of more than 2/3 of the votes within the set $\tilde{C}$ of players who emerged as the winning coalition in stage 1. In this case, player i receives a payoff of β_i, and the game ends.[12] Otherwise, stage 2 is repeated with probability

[9] With the characteristic function v given above, the Shapley value of player $i \in N$ can, for instance, be computed as $\sum_{C \subseteq S_{-i}} [(|C|!\,(n-1-|C|)!) \,/\, (n!)]\,(v\,(C \cup \{i\} - v\,(C)))$ where S_{-i} is the set of all possible coalitions that do not contain i and $n = 4$ is the number of players. The term $v\,(C \cup \{i\} - v\,(C))$ measures a player's contribution to a coalition C. For the strong player, $v\,(C \cup \{i\} - v\,(C)) = 1$ (i.e., player i is pivotal) if C consists of exactly one weak player or exactly two weak players. A weak player is pivotal if C consists of the strong player or the other two weak players.

[10] This two-stage game can be seen as a benchmark case when a coalition agreement is de facto open to renegotiation at zero cost after the formation of the coalition.

[11] Hence, the proposed split $\boldsymbol{\beta}$ satisfies $\beta_j > 0$ only if $j \in \tilde{C}$, and $\sum_{j \in \tilde{C}} \beta_j = 1$.

[12] The number of votes needed to win changes between stage 1 and stage 2. As will become clear be-

Case 1: $m_W < m_S \leq 2m_W$		Case 2: $2m_W < m_S \leq 3m_W$	
Winning coalition $\tilde{C}$	Shapley value $Sh(\tilde{C};\cdot)$	Winning coalition $\tilde{C}$	Shapley value $Sh(\tilde{C};\cdot)$
(S,W)	$\left(\frac{1}{2},\frac{1}{2}\right)$	(S,W)	$(1,0)$
(S,W,W)	$\left(\frac{2}{3},\frac{1}{6},\frac{1}{6}\right)$	(S,W,W)	$\left(\frac{2}{3},\frac{1}{6},\frac{1}{6}\right)$
(W,W,W)	$\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$	(W,W,W)	$\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$
(S,W,W,W)	$\left(\frac{1}{2},\frac{1}{6},\frac{1}{6},\frac{1}{6}\right)$	(S,W,W,W)	$\left(\frac{1}{2},\frac{1}{6},\frac{1}{6},\frac{1}{6}\right)$

Note: This table summarizes the Shapley values within all possible winning coalitions that may enter stage 2. For brevity this table does not distinguish between the identity of the weak players. A coalition (S,W,W), for instance, denotes a coalition between the strong and two weak players; there are three such coalitions denoted by (S,W,W).

Table 1: Shapley value at the time of negotiating the division of the prize.

95% and the game ends with zero payoffs for all players with probability 5%.

The dynamic nature of the prize division brings in a different logic of decision-making. Consider the second stage where a given coalition $\tilde{C}$ has to divide the prize. Table 1 summarises the set of majority coalitions and the corresponding Shapley values when the negotiation for the actual division of the prize takes place within the winning coalition $\tilde{C} \in \mathbf{C}_M$. Case 1 relates to a situation where the strong player is not "too strong"(i.e., $m_S \leq 2m_W$); case 2 represents the opposite case of $m_S > 2m_W$. A comparison of the Shapley values within (S,W) coalitions between the strong and one weak player demonstrates the potential importance of the strong player's voting rights (m_S): in a (S,W) coalition, both players have the same effective power if $m_S \leq 2m_W$ but S becomes the dominant player if $m_S > 2m_W$.[13] The Shapley values in all other majority coalitions remain independent of m_S.

low, by limiting the power of the strongest player, requiring a supermajority in stage 2 makes the stability of two-player coalitions dependent on the parameter values (m_S, m_W). With the objective of keeping the experimental setup sufficiently simple and the number of players sufficiently small (in order to reduce coordination failure in the experiment), the requirement of a supermajority within stage 2 coalitions constitutes a straightforward way of making the strong player's power meaningful. Apart from this practical reason, variations of the quota also reflect the idea that the nature of the players' interactions in the two stages need not be identical, something that may hold, for instance, when introducing explicit rules of renegotiation or when considering models of distributional conflict instead of simple voting games in order to describe the resource allocation.

[13] In this case, the strong player alone already has more than 2/3 of the votes within a (S,W) coalition and hence can dictate the allocation.

In order to compare the effect of an increase in m_S to the variant with binding agreement, suppose that all players expect that, within a winning coalition, the prize is split in stage 2 according to the players' effective voting power given by their Shapley values in Table 1. Anticipating the decisive role of the strong player when bargaining about the split of the prize, a weak player may reject an offer to form a coalition (S, W) if the strong player becomes "too strong"(case 2 in Table 1). Thus, an increase in votes and, hence, effective voting power within certain coalitions can turn into a disadvantage. Our experiment is designed to test this fundamental idea.

A straightforward formalisation of this intuitive prediction is based on behavioural assumptions on the players' strategies. Suppose that the players only consider *minimal winning coalitions*.[14] With the subset $\underline{\mathbf{C}}_M \subseteq \mathbf{C}_M$ of minimal winning coalitions being

$$\underline{\mathbf{C}}_M = \{(S, W), (W, W, W)\},$$

the strong player proposes (S, W_k) for some weak player k and must vote for a coalition (S, W_j) that has been proposed. Hence, in case of $m_S \leq 2m_W$, weak player j strictly prefers voting for (S, W_j) over voting for (W, W, W) so that the stage 1 winning coalition must be a (S, W) coalition, independent of the weak players' proposals. In case of $m_S > 2m_W$, anticipating that a weak player votes for (W, W, W) if it is proposed, the stage 1 winning coalition is (W, W, W).[15] With prize splits as in Table 1, this would imply that, in case of $m_S \leq 2m_W$, the players' ex ante expected payoffs with delayed agreement correspond to (2). But in case of $m_S > 2m_W$, the strong player's ex ante expected payoff decreases compared to her payoff in (2). In other words, the lack of commitment at the coalition formation stage harms the strong player if and only if she is "too strong."

[14] The reduced-form coalition formation game based on the payoffs in Table 1 has multiple equilibria, as is typical in coalition formation games and voting games. The fact that agreement is needed from all coalition members together with the sequential nature by which coalitions are formed makes minimal winning coalitions a natural coalition outcome since extending such a coalition can never be profitable for all members, in contrast to reducing a coalition of which a subset is winning, too.

[15] Focusing on unilateral deviations, a (S, W) coalition can also be supported as equilibrium in case of $m_S > 2m_W$ as soon as at least three different coalitions have been proposed. An example can be constructed with proposals $\{(S, W_1), (S, W_2), (W_1, W_2, W_3)\}$ if W_1 votes for (S, W_1), W_2 votes for (S, W_2) and S votes for (S, W_k), $k \in \{1, 2\}$. However, the only coalition that cannot be blocked by (a joint deviation of) a subset of players is the outcome (W, W, W) in case of $m_S > 2m_W$.

2.2 Experimental design

The experiment is based on a 2×3 between-subject design. One dimension varies the number m_S of votes of the strong player: keeping $m_W = 4$ constant, we consider treatments with $m_S = 7$ (henceforth the S_7 treatment) and treatments with $m_S = 9$ (henceforth the S_9 treatment). In both treatments, the strong player has less than 50% of the votes, but $m_S = 9$ implies that the strong player can dominate the split of the prize in stage 2 (that is, holds a supermajority) in any two-player coalition (S, W). As the second dimension, we vary the timing of agreement of the prize division and consider three different treatments: binding agreement; no ex ante binding split agreement; and no ex ante binding split agreement allowing for (non-binding) communication at the point where coalitions are proposed (i.e., before a coalition is formed in stage 1). For brevity, we refer to these three treatments as "*Binding*", "*Delayed*", and "*Delayed-Comm*" hereafter. And we will refer to the each specific experimental condition as respectively "*Binding-S_7*", "*Binding-S_9*", "*Delayed-S_7*", and so on.

2.3 Main hypotheses

Our main hypotheses consider both the effect of the binding nature of the agreement at the time of the coalition formation and of an increasing power of the strong player. Consider first the variant in which the players commit to a division of the prize at the point where they form coalitions. Hypothesis 1 states how we expect the asymmetry in the number of votes to affect the success of the strong player.

Hypothesis 1 *In the treatments with binding agreement on the split of the prize,*
(i) the likelihood to be part of the winning coalition and the average realised payoff are higher for the strong player than for a weak player, both in S_7 and S_9;
(ii) the strong player's likelihood to be part of the winning coalition and her average realised payoff are the same in S_7 and S_9.

Hypothesis 1 stems naturally from the use of the Shapley value as a way to measure the players' power in the game. In line with the asymmetries in the players' votes, the strong player as being more often the pivotal player should be part of the winning coalition more frequently than a weak player and, hence, realise a higher payoff (Hypothesis 1(i)). More importantly, since the set of majority coalitions which the players can form and the players' Shapley values are the same in S_7 and S_9, there should be no treatment effect of the increase in m_S, neither on the strong player's average payoff nor on her probability to be part of the winning coalition (Hypothesis 1(ii)).

This prediction based on the players' Shapley value can be justified further by looking at the specific sub-coalitions which are expected to be prevalent in the game. Based on non-cooperative solution concepts, any admissible coalition outcome can be supported as Nash equilibrium of the coalition-formation game.[16] However, *minimal winning coalitions* $C \in \underset{\sim}{\mathbf{C}}_M = \{(S, W), (W, W, W)\}$ are likely to be more frequent: by removing a dummy player and splitting her share between the remaining coalition members, the proposing player may increase the acceptance probability and the payoff conditional on acceptance. In the game considered, coalitions (S, W) between the strong and one weak player achieve the highest average payoff to their members, whereas symmetric coalitions (W, W, W) and splits between the three weak players may be particularly easy to agree upon and, hence, prevent coordination failure.[17] Following intuitive rather than formal arguments, we expect (S, W) coalitions to form most frequently since the strong player can prevent (W, W, W) by proposing (S, W) together with a share $\beta_W > 1/3$ whenever the corresponding weak player believes that S votes for her own proposal. Given that $\beta_W = 1/3$ can be seen as acting as a reserve price for weak players in a (S, W) coalition, we expect the strong player to be able to secure a larger share of the prize than a weak player. These considerations on possible sub-coalitions further support the first part of Hypothesis 1: strong players in *Binding-*S_7 and *Binding-*S_9 get on average higher payoffs than weak players because (S, W) coalitions are more frequent than (W, W, W) coalitions and strong players tend to get a higher payoff than weak players

[16] To illustrate, suppose that all players follow stationary strategies, let player 1 be the strong player and consider the following candidate equilibrium: Player 4 proposes $((1, 2, 4), \hat{\boldsymbol{\beta}})$ and votes for it. Players 1, 2, and 3 all propose $((1, 2, 3), \boldsymbol{\beta})$ and vote for this proposal so that $((1, 2, 3), \boldsymbol{\beta})$ wins. For players $i \in \{1, 2, 3\}$, unilateral deviations in the voting choice yield disagreement and unilateral deviations in the coalition proposal do not change the outcome (neither do unilateral deviations of player 4) so that the candidate proposals constitute an equilibrium.

[17] Even though both players are pivotal in (S, W) coalitions, the asymmetry of the players' votes may make it more difficult to agree on a split (β_S, β_W).

in (S, W) coalitions. These considerations stay the same as long as the strong player holds less than 50% of the votes, which is again mirrored in Hypothesis 1(ii).

Our second hypothesis addresses the opposite case in which there is no commitment to the prize division at the coalition formation stage. Here, we expect the bargaining power within possible winning coalitions to be important, causing a strategic disadvantage for the strong S_9 player.

Hypothesis 2 *In the treatments with delayed agreement on the split of the prize,*
(i) the strong player's likelihood to be part of the winning coalition and her average realised payoff are lower in S_9 *than in* S_7*;*
*(ii) the strong player's likelihood to be part of the winning coalition and her average realised payoff are the same as with binding agreement if she is not "too strong" (*Delayed-S_7 *compared to* Binding-S_7*), but are lower than with binding agreement if she becomes "too strong" (*Delayed-S_9 *compared to* Binding-S_9*).*

When final prize sharing decisions take place at a second stage only, the players must anticipate their power within possible winning coalitions. As with binding agreement, two-player coalitions achieve the highest average payoff for their members but must involve the strong player. Within coalitions (S, W), the weak player is equally decisive as the strong player in *Delayed-*S_7 but is dominated by the strong player in *Delayed-*S_9. In the former case of $m_S = 7$, the strong and weak players may find an agreement that makes them both (weakly) better off than under any other coalition $C \in \mathbf{C}_M$ and, hence, (S, W) coalitions may form. In the latter case of $m_S = 9$, the weak players may prefer, instead, to aim for the outcome (W, W, W). Thus, in contrast to the case with binding agreement, we expect the increase in the strong player's votes to have an effect on the coalition outcomes and to reduce her likelihood to be included in the winning coalition (Hypothesis 2(i)): moving from S_7 to S_9 should have differential effects with and without binding agreement.

Within the S_7 and S_9 treatments, we expect that the absence of binding agreement does not constitute a disadvantage in the S_7 treatments but significantly reduces the strong player's ability to form coalitions in the S_9 treatments (Hypothesis 2(ii)). Again, we can provide support for Hypothesis 2 by looking at the likely sub-coalitions to emerge in the different treatments. As argued above, if the players only consider minimal winning coalitions and expect prize splits in line with the payoffs in Table 1, a coalition (S, W) should form in *Delayed-*S_7 so that we do not expect the strong player in S_7 to be penalised from the absence of binding agreement. But the strong player should never be part of the winning coalition in *Delayed-*S_9 where the weak players are better off in the coalition (W, W, W). The latter is the payoff-dominant choice for the weak players when assuming a coalitional split of the majority coalition in line with the Shapley values in Table 1. Thus, in the S_9 treatments, the lack of commitment at the coalition formation stage should make the strong player worse off.

Third, the treatments with communication add a possibility of promising to agree on certain splits when the prize division takes place at a later stage only. From a standard theory perspective, such promises should neither affect the coalition outcomes nor average payoffs since the receivers should anticipate that the messages are not credible. But ex ante communication may affect the outcome if deviations from ex ante announcements of the intended split involve a (non-monetary) cost at least for some players, for instance due to social preferences (e.g. lying aversion, Gneezy et al. (2013)). We contrast these two alternative views on the commitment value of communication in our third hypothesis.

Hypothesis 3A *In the delayed agreement treatments with communication, the strong player's likelihood to be part of the winning coalition and her average realised payoff are the same as in the corresponding treatments without communication (*Delayed-Comm-S_7 *compared to* Delayed-S_7*;* Delayed-Comm-S_9 *compared to* Delayed-S_9*).*

Hypothesis 3B *In the delayed agreement treatments with communication,*
(i) the strong player's likelihood to be part of the coalition and her average realised payoff in S_9 *are higher than without communication (*Delayed-Comm-S_9 *compared to* Delayed-S_9*);*
(ii) the strong player's average realised payoff in S_7 *is higher than without communication (*Delayed-Comm-S_7 *compared to* Delayed-S_7*).*

Hypothesis 3A tests the standard theory prediction of no effect of communication against the alternative Hypothesis 3B based on the behavioural assumption of a non-monetary cost of deviating from ex ante announcements (for simplicity: "cost of lying"). If sufficiently many strong players exhibit a cost of lying and if this cost is sufficiently high, then communication reduces the strong S_9 player's disadvantage in the coalition formation process by restoring a possibility of limited commitment in *Delayed-Comm-*S_9. Intuitively, strong players in S_9 will promise to allocate a resource share to weak players in (S, W) coalitions which is just sufficient to make them choose the (S, W) coalition over a (W, W, W) coalition. Such an outcome can only be sustained, however, if costs of lying are widespread among strong players as well as sufficiently high to prevent ex post deviations. The alternative Hypothesis 3B(i) tests a possible positive effect of communication on strong players in S_9 against the baseline Hypothesis 3A.

The idea behind Hypothesis 3B(i) can be formalised as follows. Suppose there is a share α of players who have a cost of lying parametrised by $\frac{c_L}{2}\left(\max\{\beta_i - \mu_i, 0\}\right)^2$ where β_i is the player's prize share in the stage 2 bargaining, μ_i is the own share claimed at the communication stage, and $c_L > 0$ captures the intensity of the cost of lying. (That is, $c_L = 0$ for a share $1 - \alpha$ of players.) This is all common knowledge. Upon observing a message μ_S as part of a (S, W) proposal, the weak player anticipates her resource share at stage 2, which depends on whether she expects to face a strong player with a cost of lying (which happens with probability α) and how much the strong player's actual payoff claim would deviate from μ_S (which depends on c_L). The optimal message μ_S from the point of view of a strong player in S_9 will ensure that the weak player expects a payoff of 1/3 in a (S, W) coalition, which is what the weak player can (rationally) expect in a (W, W, W) coalition. A (pooling) equilibrium in which coalitions (S, W) form exists if and only if α and c_L are sufficiently high. If α and/or c_L is low, the weak player does not accept the coalition offer (S, W) even for very favourable (i.e., low) messages μ_S from the strong player. For details see Appendix B.

Behavioural assumptions such as a cost of lying can also change the outcome in *Delayed-Comm-*S_7. Here, the incentive to make promises is particularly important for weak players: Since the strong player can form a (S, W) coalition with any of the three weak players and since weak players may prefer a (S, W) coalition to a (W, W, W) coalition (compare the payoffs suggested by Table 1), a weak player may attempt to signal to the strong player that she would be willing to accept a resource share smaller than 1/2. Anticipating the same incentive of the other weak players, this may lead "price competition" in the promises (messages) μ_j made by weak players j, possibly down to a lower bound determined by the lowest message that is credible in the sense that a weak player with a cost of lying would still prefer the subsequent allocation in the (S, W) coalition to a (W, W, W) coalition.[18] As soon as some weak players have a cost of lying, communication on the side of the weak players obtains a commitment value and may reduce the weak players' average realised payoff in (S, W) coalitions, making the strong player better off on average (as summarised in the alternative Hypothesis 3B(ii)).

The simple model from above can also be used to demonstrate that the requirements for an effect of communication in *Delayed-Comm-*S_7 are relatively weak (see Appendix B for details). In contrast to (S, W) coalitions in *Delayed-Comm-*S_9, the strong player needs agreement from the weak player to the split in (S, W) coalitions in *Delayed-Comm-*S_7. Thus, the "risk" of believing in a promise is lower for the strong player in *Delayed-Comm-*S_7 than for a weak player in *Delayed-Comm-*S_9. Only a few players with a cost of lying (a small share α) are needed to support a (pooling) equilibrium in which all weak players promise to accept a split that favours the strong player and the strong player (rationally) expects a payoff larger than 1/2, even though only those weak players with $c_L > 0$ will indeed request less than 1/2 once the (S, W) coalition has formed.

The treatments with communication can be seen as intermediate treatments between the benchmark with full commitment (*Binding*) and the treatments without commitment (*Delayed*). Due to the random re-matching of players across rounds, repeated-game effects based on reputation and punishment are ruled out so that the strong player's disadvantage when being too strong may not fully vanish. Moreover, the size of the effect crucially depends on the credibility of the strong player's message and, hence, subsequent split behaviour. Therefore, we still expect a negative effect of the increase in votes m_S on the strong player's average payoff in the delayed agreement treatments with communication (*Delayed-Comm-*S_9 compared to *Delayed-Comm-*S_7).

[18] Such competitive effects may also be found in the treatments with binding agreement when looking at the specific sub-coalitions formed. Here, weak players can make credible split proposals that favour the strong player in a (S, W) coalition in order to be picked by the strong player (compare the discussion on sub-coalitions in the context of Hypothesis 1).

	Binding agreement	**Delayed agreement**	**Delayed agreement with communication**
S_7 ($m_S = 7$)	40 subjects 2 sessions	40 subjects 2 sessions	40 subjects 2 sessions
S_9 ($m_S = 9$)	40 subjects 2 sessions	40 subjects 2 sessions	40 subjects 2 sessions

Table 2: Overview of the 2×3 experimental design and implementation.

2.4 Experimental procedures

The experiment was run at the Centre for Behavioural and Experimental Research, Wuhan University (China) from December 2017 to May 2018. We recruited 240 participants and each joined one of 12 experimental sessions. Table 2 summarises the experimental treatments and implementations. Each session consisted of 15 rounds. At the beginning of each round, the computer randomly assigned participants into groups of four. Within a group, three players were randomly assigned to be a weak player and one to be the strong player, and each player received a number randomly drawn between 1 and 4. The numbers are used in the game to identify them and ensure anonymity.[19] Each group then played a coalitional game to determine how to divide a prize of 48 Experimental Currency Units (ECU).[20]

The game was played on computers, using the interactive software developed via the oTree platform (Chen et al., 2016). Participants were given written instructions (translated into Chinese) to read at the beginning of each session and then an experimenter also read through the instructions in front of all participants to ensure common knowledge.[21] Participants had to answer a few questions regarding the content of the instructions to check if they had understood everything correctly. After all participants completed the control questions, they were allowed to play one practice round without payment to familiarise themselves with the

[19] These choices are to prevent repeated play strategies. For the same reason, switching roles has also been used by Drouvelis et al. (2010).

[20] The amount of prize $b = 48$ instead of $b = 1$ was chosen for simplicity to rely on integers and to allow for the possibility of an equal division of the prize across all possible coalitions.

[21] Experimental screen shots and instructions (translated to English) are provided in Appendix C.

computer screens they were going to face in the main experiment. During the experiment, the sequence of actions within one round follows exactly the theory setup described above. At the end of each round, participants were informed about the final decision outcome of their own group, irrespective of whether they were part of the agreed proposal/allocation. At the conclusion of the 15 rounds, the participants were asked to complete a short set of demographic questions. Their general risk attitude was also elicited using the question proposed by Dohmen et al. (2011): "How willing are you to take risks, in general?" (answers on a 10 points Likert scale). Then, one out of the 15 rounds was randomly selected and the corresponding earnings in that round were converted to RMB at a rate of 1 ECU=2 RMB. Each session took around 1.5 hours and the average total payment including a show-up fee of 20 RMB was around 43 RMB.[22]

3 Data and Results

3.1 Data

Table 3 presents descriptive statistics on the experimental data. In total, there were 40 participants in each treatment and, hence, 150 coalitional games played. The numbers of failed negotiations (due the 5% probability for the negotiation to end after each round of unsuccessful coalition proposals) are reported in the lines of "Failed coalitions" and "Failed splits". The average number of proposals needed before a majority coalition is agreed in a game/round ranges from 1.14 to 1.47.[23] This number is fairly low, with most groups agreeing quickly. Similarly, in stage 2 of the delayed-agreement treatments, the average number of proposals needed before a split agreement is reached ranges between 1.08 to 1.75. Out of 150 coalitional games played for each treatment, few groups failed to reach an agreement before

[22]A typical hourly wage for a student research assistant is around 30 RMB at that time.

[23]Figure 2 in Appendix A.5.1 provides distributions of the number of proposals required to form a winning coalition in each treatment. In most cases, agreements were reached immediately within the first proposal, especially in later rounds (8-15) where more than 90% of the cases agreement was reached based on the initial proposals. The data also suggests that communication helps to slightly reduce the average number of proposals that are needed to form winning coalitions in both S_7 and S_9 treatments, which is mostly because communication facilitates the formation of (S, W) coalitions (see Tables 13 and 14 in Appendix A.5.2).

	Binding agreement		Delayed agreement		Delayed agreement w/comm	
	S_7 (7,4,4,4)	S_9 (9,4,4,4)	S_7 (7,4,4,4)	S_9 (9,4,4,4)	S_7 (7,4,4,4)	S_9 (9,4,4,4)
Num.subjects	40	40	40	40	40	40
Num.games	150	150	150	150	150	150
Avg.s1.prop/round	1.36	1.47	1.38	1.39	1.14	1.20
Avg.s2.prop/round	-	-	1.75	1.41	1.39	1.08
Failed coalitions	3	5	3	4	0	3
Failed splits	-	-	9	2	4	4

Note: "Avg.s1.prop/round" and "Avg.s2.prop/round" respectively indicate the average number of proposals required in stage 1 and stage 2 until an agreement is reached. When no agreement is reached and the game is randomly terminated by the computer, it is indicated as failed coalition (if it happened in stage 1) or failed split (if it happened in stage 2). In the binding agreement condition the failed coalitions are also failed splits given that there is no stage 2.

Table 3: Descriptive Statistics.

the random discontinuation (see the last two rows of Table 3).[24]

3.2 Main results

Figure 1 provides an overview of the types of majority coalitions formed (as share of the total number of coalitions formed) in each treatment. In the S_7 treatments where the strong player is not "too strong", the lack of commitment at the coalition formation stage does not affect the coalition outcome. In all three variants of the S_7 treatment, we mostly observe (S, W) coalitions among the winning coalitions: 86.4% in *Binding-*S_7, 89.1% in *Delayed-*S_7, and 90.7% in *Delayed-Comm-*S_7.[25]

This finding on the role of commitment drastically changes in the S_9 treatments (see Figure 1) where the strong player becomes "too strong". The possibility of binding agreement compensates for the increase in strength (votes) so that the share of (S, W) coalitions stays high in *Binding-*S_9 (it even slightly increases to 94.5%). With delayed agreement, however, this share drops to 37% and 35.4% in the *Delayed-*S_9 and *Delayed-Comm-*S_9 treatments

[24]Our subsequent analysis will focus on the type of coalitions formed. We provide a summary of the initial coalitions proposed by the weak and the strong players in Tables 9 and 10 in Appendix A.2.

[25]In most other cases in all the S_7 treatments, the coalition outcome is (W, W, W); see also the summary Table 8 in the Appendix A.1.

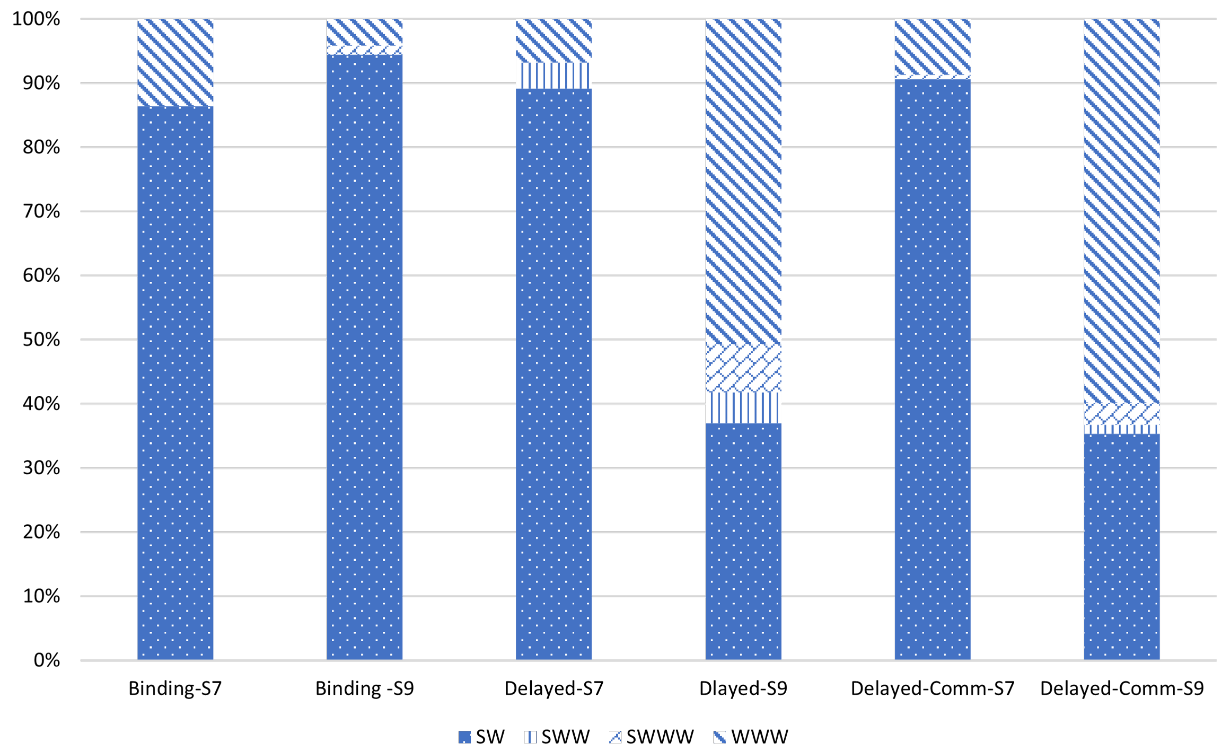

Figure 1: Percentages of winning coalitions by treatment and coalition type. The minimal coalitions (SW and WWW) form the great majority of coalitions formed.

respectively. In turn, (W, W, W) coalitions without the strong player now become the most frequent coalition outcome (50.7% and 59.9% respectively). As seen in Figure 1 and in line with our prediction, non-minimal coalitions are rarely formed.

To assess the significance of the differences across conditions, we estimated a multilevel mixed regression model:

$$y_{jik} = \alpha + \sum \beta_j treatment_j + session_i + participant_k, \tag{3}$$

where y_{jik} is the outcome of interest, either a binary variable coding for being in the winning coalition or not or the proportion of the payoffs secured, α is a constant, the variable $treatment_j$ are dummy variables coding for all the possible treatment conditions j displayed in Table 2. The variable $session_i$ is a random effect for session i and $participant_k$ is a random effect for participant k. We first estimated this model for weak and strong players separately and then we estimated a model interacting each treatment with the strength of the player to measure the significance of the difference between weak and strong players. Table 4 presents the results of the corresponding estimations. The first two pairs of columns present the results for strong and weak players separately and the last two columns present the difference between them.[26]

We find that Hypothesis 1 on the advantage of the strong player in case of binding agreement is supported by the data. First, the strong player is significantly more likely to be part of the winning coalition than a weak player: 86% vs 42% in the *Binding*-S_7 treatment (difference: 44%, $p < 0.001$) and this difference further increases by 15% ($p = 0.014$) in the *Binding*-S_9 treatment. Consequently, the strong players are able to secure larger payoffs than weak players: on average 57% vs 14% in the *Binding*-S_7 treatment (difference: 43%, $p < 0.001$) and this difference further increases by 9% ($p < 0.001$) in the *Binding*-S_9 treatment. These results align with part (i) of Hypothesis 1. Second, as shown by the coefficients of "Binding-S_9" in the first column, there is only a moderate and insignificant effect of an increase in the strong player's voting rights on the player's chances to be in the winning coalition ($p = 0.304$) and on her share of the payoff ($p = 0.246$) when agreement is binding.

[26]We also estimated a corresponding multilevel logistic model that shows similar results.

	Probability to be in a winning coalition					
	Strong Players		Weak Players		Difference	
	Coef.	Std.Err.	Coef.	Std.Err.	Coef.	Std.Err.
Reference treatment:						
Binding-S_7 (base)	0.86***	0.06	0.42***	0.04	0.44***	0.04
Differences (to base):						
Binding-S_9	+0.09	0.09	-0.05	0.05	0.15*	0.06
Delayed-S_7	+0.07	0.09	-0.03	0.05	0.10	0.06
Delayed-S_9	-0.37***	0.09	+0.31***	0.05	-0.68***	0.06
Delayed-Comm-S_7	+0.05	0.09	-0.03	0.05	0.08	0.06
Delayed-Comm-S_9	-0.46***	0.09	+0.33***	0.05	-0.79***	0.06
	Average payoff					
	Strong Players		Weak Players		Difference	
	Coef.	Std.Err.	Coef.	Std.Err.	Coef.	Std.Err.
Reference treatment:						
Binding-S_7 (base)	0.57***	0.04	0.14***	0.01	0.43**	0.02
Differences (to base):						
Binding-S_9	+0.07	0.05	-0.02	0.02	0.09***	0.03
Delayed-S_7	-0.10	0.06	+0.03	0.02	-0.14***	0.03
Delayed-S_9	-0.25***	0.06	+0.08***	0.02	-0.34***	0.03
Delayed-Comm-S_7	-0.01	0.06	0.00	0.02	-0.03	0.03
Delayed-Comm-S_9	-0.31***	0.06	+0.10***	0.02	-0.42***	0.03

Note: Estimation by multilevel mixed model, with session and participant random effects. The coefficients are tested against zero and the significance levels are marked as *** 0.1%, ** 1%, * 5%.

Table 4: Treatment effects on players' probability to be in the winning coalition and on their average payoffs.

These results align with part (ii) of Hypothesis 1.

Result 1 (Binding agreement) *In the* Binding *treatments, the strong player is more likely than a weak player to be in the winning coalition and is able to secure a larger payoff. Moreover, strong players in treatments S_7 and S_9 do not significantly differ in their chance to be in the winning coalition and in their average payoff.*

The absence of a binding agreement at the time of the coalition formation does not change the picture when the strong player is not too strong as in the *Delayed-S_7* treatment. As already suggested by Figure 1, however, this changes markedly when the strong player becomes too strong as in the *Delayed-S_9* treatment where her voting power at stage 1 is counterbalanced by the low bargaining power of weaker players at the second stage. In line with Hypothesis 2(i), a strong player is significantly less likely to be in the winning coalition in *Delayed-S_9* than in *Delayed-S_7* (49% compared to 93%).[27] The strong player's payoff is also significantly lower in *Delayed-S_9* (32%) than in *Delayed-S_7* (47%).[28] Moreover, the difference between a strong player in *Delayed-S_7* and *Binding-S_7*–both in terms of her ability to form a winning coalition and in terms of the average payoff–is not significant ($p = 0.461$ and $p = 0.114$ respectively; see also the coefficients of "Delayed-S_7" in the first column of the upper and lower part of the Table 4). In the S_9 treatments, however, the strong player's likelihood to be in the winning coalition and her average payoff decrease if agreement is delayed (by 46% and 32%, respectively; compare the difference between coefficients of *Binding-S_9* and *Delayed-S_9* in Table 4, $p < 0.001$ in both cases.). The latter results align with Hypothesis 2(ii). The negative impact of the absence of a binding split agreement for the strong player in *Delayed-S_9* is big enough such that the chance to be in the winning coalition is even lower for the strong player than for a weak player (49% vs. 73%; $p < 0.001$).[29]

Result 2 (Without binding split agreement) *In the* Delayed-S_9 *treatment, the strong*

[27] The difference between these two numbers, −44%, corresponds to difference between the coefficients of "Delayed-S_9" and "Delayed-S_7" in the upper part of Table 4; $p < 0.001$.

[28] The difference between these two numbers, −15%, corresponds to the difference between the coefficients of "Delayed-S_9" and "Delayed-S_7" in the lower part Table 4; $p < 0.014$.

[29] At the same time, the strong player's average payoff decreases to 32%, while a weak player's payoff increases to 22% in the *Delayed-S_9* treatment.

player's chances of being in the winning coalition and her average payoff are reduced relative to a strong player in either Delayed-S_7 *or the* Binding *treatments. There is no such negative effect for a strong player in* Delayed-S_7 *relative to the* Binding *treatments.*[30]

Finally, consider the effect of communication at the coalition formation stage. We test the standard theory prediction of no effect of (non-binding) ex ante communication (Hypothesis 3A) against the alternative Hypothesis 3B that communication restores some form of limited commitment. The alternative hypothesis is based on the idea that (some) players have a non-monetary cost of deviating from a prize sharing intention communicated ex ante and that the receivers of such a message anticipate the implicit commitment value for (some) players.

Hypothesis 3B(i) conjectured that communication should help the strong player in *Delayed-Comm-S_9* to attenuate the risk perceived by weak players when joining a coalition with them, thereby mitigating the "too strong" player's disadvantage caused by the absence of binding agreement. Our results do not support this conjecture but rather provide support for Hypothesis 3A. First, the strong player is slightly less likely to be in the winning coalition in the *Delayed-Comm-S_9* than in *Delayed-9* (40% vs. 49%), although the difference is not significant ($p = 0.334$). Consequently, the strong player's average payoff is also slightly lower in *Delayed-Comm-S_9* than in *Delayed-S_9* (26% vs. 32%); this difference is not significant either ($p = 0.325$). Hypothesis 3B(ii) conjectured that communication in *Delayed-Comm-S_7* may lead to competition among the weak players for being selected by the strong player and, hence, increase the strong player's payoff in (S, W) coalitions. On average, the strong player's average payoff is 9% higher in *Delayed-Comm-S_7* than in *Delayed-S_7*, but the difference is not statistically significant ($p = 0.151$).[31]

Result 3 (With communication) *When the split agreement is delayed, the possibility of communication during the coalition negotiation does not improve the "too strong" player's*

[30] To further investigate the weak players' preference for (against) including the strong player, Table 16 in Appendix A.7 shows that the treatment effect on a weak player's likelihood to include the strong player in her initial coalition proposal is robust to controlling for individual-specific characteristics such as risk preferences, gender, etc. This finding supports the importance of strategic considerations about bargaining power generated by the treatment variations.

[31] See the difference between the two coefficients of *Delayed-Comm-S_7* and *Delayed-S_7* in the lower part Table 4; the p-value is obtained through a Wald test.

chance to be in the winning coalition and her average payoff in Delayed-Comm-S_9 *as compared to* Delayed-S_9. *Similarly, communication does not significantly increase the strong player's average payoff in* Delayed-Comm-S_7 *as compared to* Delayed-S_7.

3.3 Balancing or bandwagoning

Beyond our three main hypotheses, Figure 1 on coalition outcomes and Table 11 on average realised payoffs (in Appendix A.3) reveal interesting patterns in regard to how the game's outcomes changed across treatments. As already discussed above, when moving from binding agreement to delayed agreement in S_9, the proportion of (S, W) coalitions drops markedly (from 94.5% to 37%). However, it remains substantial even in *Delayed-*S_9 (37%), in spite of the decisiveness of the strong player in the second stage negotiation within this two-player coalition. This higher-than-predicted share of two-player coalitions with the strong player in *Delayed-*S_9 is consistent with the average payoffs weak players received in such coalitions (26%) (see Table 5 below).[32] In this section, we discuss the players' payoffs across different coalition types in more detail.

The previous results provide clear evidence in favour of balancing strategies when weak players are considering forming a coalition with or without a strong player in the *Delayed-*S_9 treatment. They also suggest that balancing–i.e., a coalition (W, W, W)–is not always the best option for a weak player: bandwagoning with the strong player–i.e., a coalition (S, W)–may be advantageous if the risk of joining a strong player is limited as in the *Delayed-*S_7 treatment. Intuitively, Table 1 shows that a weak player may actually prefer a (S, W) coalition with a strong player in *Delayed-*S_7 treatment, as both have equal bargaining power when dividing the prize in the winning coalition.

The experimental data confirm this intuition. Table 5 presents the average payoff of a weak player in the different coalitions for each treatment. In the (S, W) coalitions in *Delayed-*S_7, a weak player secures an average payoff close to one half of the prize (0.49). Here, bandwagoning with the strong player turns out to be the best option for a weak player. Indeed, while bandwagoning decreases markedly between *Binding-*S_9 (94.5%) and

[32] In only 43% of the (S, W) coalitions formed in the *Delayed-*S_9 treatment, the strong player offered zero to the weak player.

	Binding agreement		Delayed agreement		Delayed agreement with communication	
	S_7 (7444)	S_9 (9444)	S_7 (7444)	S_9 (9444)	S_7 (7444)	S_9 (9444)
SW	0.34	0.32	0.49	0.26	0.38	0.29
SWW	-	-	0.19	0.16	-	0.25
WWW	0.33	0.33	0.33	0.33	0.33	0.33
SWWW	-	0.25	-	0.25	0.25	0.24

Table 5: Average payoff for a weak player in each winning coalition type (as a share of the prize).

Delayed-S_9 (37%), it, on the contrary, slightly increases between *Binding-*S_7 (86.4%) and *Delayed-*S_7 (89.1%). In the latter case, whereas in *Binding-*S_7 a bandwagoning coalition (S, W) only results in a slightly higher payoff to the weak player (0.34 as compared to 0.33 in balancing coalitions (W, W, W)), bandwagoning becomes clearly more profitable for weak players in *Delayed-*S_7. In *Delayed-*S_9, however, bandwagoning coalitions yield lower average payoff to weak players (0.26) than balancing coalitions (0.33) even though the weak player's payoff in (S, W) coalitions is still considerably higher than predicted (in the line with the higher-than-predicted share of (S, W)-coalitions that form in *Delayed-*S_9). Overall, in the *Delayed* treatment where the players must form expectations on the split agreements in different sub-coalitions, the players' voting behaviour at the coalition formation stage is consistent with the payoffs they can rationally expect in the possible coalitions.[33]

Table 5 also shows that communication only has weak effects on payoffs in the S_9 treatments: the weak player's payoff in (S, W) coalitions only slightly goes up (from 0.26 to 0.29), in line with Result 3 and Hypothesis 3A. However, communication significantly reduces the weak player's payoff in (S, W) coalitions of the delayed S_7 treatments (average 0.49 vs 0.38, $p < 0.001$), which is consistent with the intuition of competition for (S, W) coalitions among the weak players (Hypothesis 3B(ii)). This, in turn, increases the strong player's payoff in (S, W) coalitions, but the increase is not strong enough for communication to significantly

[33] We also explored potential history effects of coalition proposals on weak players' payoff in (S, W) coalitions of the treatments with delayed agreement. Table 16 (in Appendix A.7) shows that the fact that a (W, W, W) coalition had been proposed and, hence, could have been chosen in stage 1 as well does not change the weak player's payoff if she instead formed a (S, W) coalition. Whether or not the weak player herself proposed the coalition has no effect on the second stage split either.

	All coalition types		
	Weak	Strong	Total
Delayed-Comm-S_7	24% (126)	61%(62)	36%(188)
Delayed-Comm-S_9	3%(161)	56%(41)	14%(202)
Total	12%(287)	59%(103)	24%(390)
	(S, W) coalitions		
	Weak	Strong	Total
Delayed-Comm-S_7	29% (103)	61%(62)	41%(165)
Delayed-Comm-S_9	19%(27)	55%(38)	40%(65)
Total	27%(130)	59%(100)	41%(230)

Table 6: Untrustworthy messages sent with the coalition proposals in rounds when winning coalitions are formed, by treatment and strength.

increase the strong player's average payoff across all coalition outcomes in the S_7 treatments.

3.4 The role of communication

The treatments with communication provide interesting data on the role of communication in the negotiation process and its effect on final outcomes. In this section, we turn to an analysis of the messages sent in the treatments with communication.

Hypotheses 3A and 3B address the question of whether ex ante messages can credibly transmit intentions to split and, hence, affect the coalition outcome. We coded the set of messages to extract the concrete proposals of splits which the text messages may contain. The coding identified whether messages were sent or not, whether the message contained a split proposal and, if so, the amounts proposed. Our analysis focuses on the messages sent when coalition formation was successful (recall that the coalition formation stage was repeated if no agreement was reached). Using the messages of those players who ended up to be included in the winning coalition, there were 718 such messages in total from the two treatments with communication. Among these messages, 390 included a (non-binding) split proposal (see Table 15 in Appendix A.6 for a summary of the type of messages sent).

Table 6 compares the (non-binding) split proposal indicated through communication at the stage of coalition formation to the actual split proposed after the coalition has formed in order to identify whether the winning coalition members' announcements of their intended split are trustworthy or not. We define a message as untrustworthy if the actual stage 2 proposal is lower than the intended split announced ex ante in the message. Accordingly, Table 6 shows the proportion of messages which were untrustworthy as a function of the treatment and the strength of the player (with the base number provided in brackets). The upper part of the table presents the data for all the coalition types and the lower part presents the data for (S, W) coalitions only; close to two thirds of the messages with a split proposal are sent in the latter (230/390). We observe that the majority of the messages from strong players are untrustworthy. This likely explains why communication does not help strong players to convince weak players to join them in a coalition, that is, why Hypothesis 3B(i) is rejected: a large proportion of strong players use their message to entice weak players in a coalition but do not follow on their initially stated intentions. If this is correctly anticipated (or experienced) by the weak players, there is no room for a strong player to benefit from the possibility of ex ante communication on split intentions.[34]

We also observe that a substantial proportion of weak players send untrustworthy messages in (S, W) coalitions. This is particularly the case in the *Delayed-Comm-*S_7 treatment (29%) where a weak player has a veto power within the coalition in the second stage. There is therefore an incentive to state overly modest payoff claims in the first stage and revise these claims upward in the second stage once the coalition has formed. In contrast to the scenario for weak players in *Delayed-Comm-*S_9, however, strong players should opt for (S, W) even when they anticipate a general untrustworthiness of the weak players. (They could nevertheless expect to get at least 50% of the prize.) Thus, as soon as some (possibly few) weak players exhibit a non-monetary cost of deviating from the intention to split, this would increase the strong player's average payoff in (S, W) coalitions (as captured by Hypothesis 3B(ii)).

Table 7 compares the minimum prize share a player is willing to accept based on her message to the actual split proposal in stage 2, as a function of treatment (S_7 or S_9) and player's strength (weak or strong). In all four cases, the amount requested in stage 2 is on average higher than the split proposal indicated in the ex ante message in stage 1. For strong

[34]Along the lines of the theory, the share α of players with a "cost of lying" and/or the parameter c_L which measures the intensity of the cost of lying are low empirically.

	Weak		Strong	
	message	actual-proposal	message	actual-proposal
Delayed-Comm-S_7	34%	39%	55%	63%
Delayed-Comm-S_9	38%	40%	49%	70%

Table 7: Minimum willingness to accept in messages vs. in actual split-proposals, by treatment and strength

players, the actual proposal is on average 8% higher in *Delayed-Comm-*S_7 and 21% higher in *Delayed-Comm-*S_9. This pattern helps further understand why strong players do not benefit from communication in *Delayed-Comm-*S_9 relative to *Delayed-*S_9: the strong players in S_9 tried to lure weak players into a coalition agreement they often had no intention to respect. Belief formation and updating of the weak players would make the ex ante messages ineffective, in line with the results on coalition outcomes.

With respect to the weak players' messages, Table 7 shows that, in *Delayed-Comm-*S_7, the share that weak players demand–both in their message and in their actual proposal (34% and 39%, respectively)–is lower than what they get in *Delayed-*S_7 without communication (49%; compare Table 5). This supports the idea that the possibility to send messages to the strong player intensifies the competition between the weak players who propose more attractive splits to the strong player and, at least on average, also claim a lower payoff in stage 2 than without ex ante messages.

4 Discussion and applications

The strategic considerations analysed in this stylised experiment may be reflected in incentives for coalition formation at the political level and in international relations.

4.1 Government coalitions

Our results give news insights into the logic of multi-party coalition governments. When splitting rewards within a government coalition, the widely supported "Gamson law" asserts that the division of portfolio tends to be close to the parties' vote shares (Gamson, 1961). However, one noticeable deviation from this predication has been observed: the "small party bias" whereby small parties get more portfolios than their share of votes in coalitions with few members (Browne and Franklin, 1973). This phenomenon has been considered puzzling, since Baron and Ferejohn (1987)'s model of legislative bargaining predicts the opposite: the party in charge of forming a government coalition (typically the largest in terms of seats) can extract a rent by its position as proposer. In contrast to this prediction, the considerations about the non-binding nature of government coalitions can help explain the small-party bias. A small party can extract large benefits from being in a coalition with a big party. Coalition agreements are not binding by nature and parties are able to drop from a coalition at any time. As a consequence, the big party is unable to secure the future support of the small party. The risk of defection from the small party, which could entail the fall of the government gives it a bargaining power disproportionate to its share of members of parliament.[35]

This mechanism also explains why the small party bias has not been found in large coalitions with many parties (Browne and Franklin, 1973). In such coalitions, any given small party is less likely to provide critical support without which the government may lose the required majority.

4.2 Coalition of regional entities as a country

In human history, the shaping of countries' boundaries is often the result of violent conflicts ending in the acquisition or secession of territories. However, there are also instances where the inclusion of a territory is the result of a non-violent agreement. The decision for a regional entity to be part of a larger country fall into two categories: the decision of independent entities to join another country and the decision of a part of a country to secede.

Joining a country Choices to join another country are particularly interesting to analyse when being decided by referendum.[36] Such situations most often emerged after wars where the settlement on the final borders was agreed to involve the consent of local populations.[37]

[35]We observe this small player advantage in our experiment. In the *Delayed-*S_7 treatment, where a weak player has *de facto* a veto power within the coalition, the weak player is able to secure nearly half the prize.

[36]When elites are deciding, there is the possibility that their incentives are not aligned with the long term interest of the population as they may extract some personal rent from their decisions (e.g. a larger country can buy their support).

[37]The most notable examples are the 1920 Carinthian referendum (outcome: decision becomes part of Austria instead of Hungary), the 1920 Schleswig referendum (outcome: division of the Schleswig between

Ethnic and cultural proximity or differences typically play a major role in these decisions, as well as economic advantages and disadvantages. We contend here that the commitment value of formal agreements is reflected in such decisions, too. Our study puts forward two insights relevant to this question: First, a union between similarly sized (similarly powerful) entities is likely more attractive for a (typically) small territory than a union with an asymmetrically large territory. Second, a country able, by design, to commit to binding agreements is likely to be a more attractive option to join.

Two iconic examples may illustrate the importance of these considerations. Puerto Rico which aims to become a US state; and Taiwan which rejects unification with China. Puerto Rico, a small islands of the Caribbean, was invaded by the USA in 1898 during the Spanish-American war. In spite of ethnic, cultural and linguistic differences, Puerto Ricans rejected independence in repeated referenda (1950, 1967, 1993, 1998) and in 2012 and 2017 they voted in favour of statehood. In contrast to this situation Taiwanese share common linguistic and cultural heritage with mainland China. The country is however de facto independent and resists pressure from Beijing towards unification.

At first sight, the two situations places a small territory in the position of accepting or rejecting to be part of a large and powerful country. But the political prospects associated with unification are arguably markedly different. The USA is a federal country composed of states. Two of them, Rhode Island and Delaware, are smaller than Puerto Rico. States have a large autonomy to organise local affairs and their rights and relationship between each other and towards the federal government is guaranteed by the US Constitution which has been relatively unchanged since 1788 (only 27 amendments have been made). In that sense, joining the USA is more similar to joining a set of states where small states' rights are protected by the Constitution which makes any union agreement credibly binding. Certainly as a consequence, even though a specific Puerto Rican national identity has grown over the

Denmark and Germany), the 1921 Sopron referendum (outcome: stayed in Hungary instead of joining Austria), the 1935 and 1955 Saar referenda (outcome: staying in Germany instead of joining France), the 1947 North West Frontier province's and Sylhet's referenda (outcome: the provinces choose to belong to Pakistan instead of India), the 1961 British Cameroons referendum (outcome: division of Northern and Southern Cameroon between Nigeria and Cameroon), the 1967 Gibraltar referendum (outcome: staying part of the UK instead of joining Spain), the 1976 referendum in Mayotte (outcome: staying part of France instead of joining the Comoros), the 1986 and 2013 referenda in the Falklands (outcome: staying part of the UK instead of joining Argentina).

years, Puerto Rican appreciate the opportunity of having the US citizenship (Duany and Pantojas-Garcia, 2006), and majoritarily support for the territory to become a full part of the USA as a state.

In comparison, China does not have a federal model of government but a centralised one with tremendous prerogatives in Beijing to rule and decide how local matters should be settled. The authoritarian nature of the regime limits its ability to credibly commit to some rules of agreement. In 1981, the Chairman of the NPC's Standing Commitee, Ye Jiangying articulated the guidelines to return Taiwan to mainland China. He promised after an eventual reunification a special administrative region (SAR) status with a "high degree of autonomy" which could retain its armed forces while Beijing would not interfere in its internal system. But, in its attempt at a reunification, mainland China is hindered by a credibility problem in regard to the long term validity of a commitment to local autonomy and possible renegotiation. The key insight we put forward is that this asymmetry in a future union and the inability to enshrine rights for the local authority in a binding agreement contributes to making the prospect of unification with China unappealing for Taiwan, in spite of the cultural proximity and likely economic gains from closer ties with mainland China.[38]

Seceding from a country Similar strategic considerations can underlie the decision of a territory to secede or not from an existing country. Here again, while the contexts of every situation differ in many ways, we believe that the insights we discuss in this study help shed light on the preferences for independence.

To the extent that small territories have the possibility to opt out from a country, we can think of the choice to stay in or to move out as a choice to be part of coalition with other territorial areas. The two insights we have put forward are the following: small partners will find a coalition less appealing if it contains an oversized partner and if it does not offer a binding agreement on resource distributions. Countries are characterised by more of less autonomy devolved to its constitutive regions. At one end, federal countries grant a lot of legislative and judiciary autonomy and at the other end centralised countries give minimum autonomy to local regions. As a result, by limiting the power of the central government,

[38]As described by Tubilewicz (2016), Taiwanese are "largely reluctant to hand over their hard won democratic and civil liberties to Beijing in the name of unity with an authoritarian entity".

federal countries offer more guarantee to local jurisdictions about future decisions than a centralised country.

The role of governance structures on secessionists aspirations is well acknowledged in political science. It is generally thought that greater decentralisation of power (for example, through federalist structures) "avert centre-periphery ethnic conflict, or secession" (World Bank, 2011, p. 166). Such solutions have often been pushed by international organisations to resolve intra-national conflicts (Wolff et al., 2020). Despite concerns that greater decentralisation may foster local secessionist movements, the evidence seems to suggest that in federal democracies, successful secession are extremely rare and that it is often the refusal to decentralise power which foster local support for secession (Bermeo, 2002).

As an implication of our study, we can expect secessionists preferences to be weaker in situations where federal governance offers a credible guarantee of respect of future autonomy than in situations of unilateral agreement with a powerful centre. Interestingly, Rode et al. (2018) found that federal structures which are symmetric (giving identical status to the different territorial entities constituting the country) are more effective at reducing secessionists' preferences than federal structures which are asymmetric and treat a specific territory as an exception. The second situation typically happens in relationships between a powerful centre and a smaller periphery and the long term credibility of such ad hoc institutional agreements is arguably relatively weaker.

Some well known cases can illustrate these general considerations: Scotland within the UK, Catalonia within Spain, and Quebec within Canada. Spain and the UK are not fully centralised country as they grant autonomy to sub-areas: autonomous regions in Spain, the "Home Nations" in the UK. But these two countries are not fully federal either and still allocate more decision power to the central government. Both the UK and Spain can be seen as articulated around one large geographical entity where the countries capital is located: England in the UK and the Spanish (Castillan) speaking regions in Spain. As a consequence, both Scotland and Catalonia are smaller partners in a country. This centre-periphery pattern is also articulated around an asymmetric model of devolvement where Scotland and Catalonia are given specific levels of autonomy relative to the rest of the country.[39] With the growing

[39]The clearest example is the fact that all Home Nations have devolved parliaments, except England which

integration of the European Union, the option of becoming a smaller state with equal rights in a broader union (Europe), rather than a periphery in an asymmetric union (UK, Spain) has seemingly played a positive role in Scotland's and Catalonia's secessionist movements (Bourne, 2014).[40]

Quebec in Canada presents a different case. Even more than in Spain, French speaking Quebec appears like a smaller geographical partner in a union of states with a different language. Initially colonised by France and conquered by England, the territory saw failed attempts at independence, and requests for more autonomy emerged after the Second World War. The frustration of these requests led to two referenda (1980, 1995), with the second one failing only by whisker in 1995 (50.5% for the "No"). While the result may very well have been different, the calls for independence have substantially decreased since. In that process, the federal nature of Canada can be seen as an advantage. It protects the autonomy of local decisions more strongly.[41] These characteristics make Canada more appealing as a partnership of geographical entities. And unlike in Europe, there is no alternative union present as a type of exit option.

4.3 Coalitions of countries in alliances

At the international level, the idea that countries may form alliance to counteract the strength of a more powerful country is at the heart of the notion of "balance of power" which can be seen throughout European history (Hume, 1994). The modern theory of international relations places this strategy alongside bandwagoning, that is, siding with the powerful country (Waltz, 2010). Concerns about the binding nature of alliances help understand conditions which favour a choice towards balancing or bandwagoning.

As an illustrative example, Russia's inability to convince former soviet countries to join an economic union under its umbrella can be interpreted as a "too big to prevail" situation. After the fall of the Soviet Union, Russia tried to foster an economic association with other post-soviet states with the Community of Independent States (CIS). But this attempt has

constitutes the powerful centre of the union.

[40] As illustrated by the slogan "independence in Europe" used by the Scottish National Party.

[41] As a reflection of the absence of centre-periphery patter, the federal capital is, on purpose, a medium size city, Ottawa, rather than a historical centre of power.

mainly been a failure with other states resisting deeper integration and often looking for alternative alliances instead (Kubicek, 2009). This may seem surprising a priori given the economic, cultural and linguistic bonds between these different countries and between these countries and Russia. But the history of the CIS illustrates the concerns of tying oneself in an alliance with a hegemonic partner. Very early on, Russia used pressure to force the participation of some states (Georgia, Moldova) and stated explicitly, via the so-called "Monroeski Doctrine", its right to intervene in CIS states to protect the right of ethnic Russians. CIS skeptics countries formed a specific alliance without Russia, the GUAAM group (Georgia, Ukraine, Uzbekistan, Azerbaijan, Moldova).[42] This choice is strikingly evocative of an alliance of small players in our experiment with the goal of counter balancing the strength of a potential hegemonic coalition partner. One of the challenges Russia faces in its endeavour to convince other states to join its alliance is a credible commitment to a long term governance agreement. The lack of democratic rule of law at home and a record of military and hybrid interventions in other countries undermines Russia's credibility about the guarantee of future relationships.

By comparison, the European Union is, if anything, politically weak and ineffective internationally. But as a coalition of countries it does not suffer from the two issues faced by the CIS. The European Union credibly guarantees the respect of the sovereignty of its members and it is composed of middle-sized and small-sized countries with which a new country can build flexible coalitions without risking being systematically isolated. These characteristics offer to new members the prospect of participating on an equal footing in future intra-bloc negotiations. It likely explain why, in spite of its weaknesses, the European Union has acted as a magnet for Eastern European countries (Vachudova and Hooghe, 2009).

5 Conclusion

In this study we investigated the effect of players' power on the formation of coalitions to divide a prize when the assumption that agreements are binding is relaxed. In an experimental weighted majority game we find that a player's strength in terms of voting rights

[42] Which became GUAM with the departure of Uzbekistan in 2005

predicts the players' chances to be in the winning coalition when the coalition agreement is binding. However, strength is not necessarily an advantage when binding agreements about how to share the coalition's prize are not possible at the point where the coalition is formed. Without binding agreements, the players' strength which make them pivotal when forming coalitions can also make them too powerful within the winning coalition at the time where the actual split of the prize will have to be agreed upon. Weak players therefore opt to switch to coalitions characterised by a balance of power between coalition members rather than allying with a "too strong" player.

We discussed how these strategic implications of within coalition bargaining likely explains key features of coalitional bargaining in a wide range of real world situations such as government coalitions, federal unions and geopolitical alliances.

References

Acemoglu, D., Egorov, G., and Sonin, K. (2008). Coalition formation in non-democracies. *The Review of Economic Studies*, 75(4):987–1009.

Agranov, M. and Tergiman, C. (2014). Communication in multilateral bargaining. *Journal of Public Economics*, 118:75–85.

Baranski, A. and Kagel, J. H. (2015). Communication in legislative bargaining. *Journal of the Economic Science Association*, 1(1):59–71.

Baron, D. P. and Diermeier, D. (2001). Elections, governments, and parliaments in proportional representation systems. *Quarterly Journal of Economics*, 116(3):933–967.

Baron, D. P. and Ferejohn, J. (1987). Bargaining and agenda formation in legislatures. *The American Economic Review*, 77(2):303–309.

Bermeo, N. (2002). A new look at federalism: The import of institutions. *Journal of Democracy*, 13(2):96–110.

Bloch, F. and Dutta, B. (2011). Formation of networks and coalitions. In Benhabib, J., Bisin, A., and Jackson, M. O., editors, *Handbook of Social Economics*, volume 1, chapter 16, pages 729–779. Elsevier.

Bourne, A. K. (2014). Europeanization and secession: The cases of catalonia and scotland. *JEMIE*, 13:94.

Browne, E. C. and Franklin, M. N. (1973). Aspects of coalition payoffs in european parliamentary democracies. *American Political Science Review*, 67(2):453–469.

Chen, D. L., Schonger, M., and Wickens, C. (2016). oTree–an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.

Diermeier, D. (2011). Coalition experiments. *Cambridge Handbook of Experimental Political Science*, page 399.

Diermeier, D., Swaab, R. I., Medvec, V. H., and Kern, M. C. (2008). The micro-dynamics of coalition formation. *Political Research Quarterly*, 61(3):484–501.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the european economic association*, 9(3):522–550.

Drouvelis, M., Montero, M., and Sefton, M. (2010). Gaining power through enlargement: Strategic foundations and experimental evidence. *Games and Economic Behavior*, 69(2):274–292.

Duany, J. and Pantojas-Garcia, E. (2006). Fifty years of commonwealth: The contradictions of free. *The Okinawan Journal of American Studies*, 3:7–27.

Farquharson, R. (1969). *Theory of voting*. Blackwell.

Felsenthal, D. S., Machover, M., et al. (1998). The measurement of voting power. *Books*.

Fiorina, M. P. and Plott, C. R. (1978). Committee decisions under majority rule: An experimental study. *American Political Science Review*, 72(2):575–598.

Frechette, G., Kagel, J. H., and Morelli, M. (2005). Behavioral identification in coalitional bargaining: An experimental analysis of demand bargaining and alternating offers. *Econometrica*, 73(6):1893–1937.

Gamson, W. A. (1961). A theory of coalition formation. *American Sociological Review*, 26(3):373–382.

Geller, C. and Mustard, J. (2007). Focused power: Experimental manifestation of the shapley-shubik power index. *Available at SSRN 1052521*.

Gelman, A., Katz, J. N., and Bafumi, J. (2004). Standard voting power indexes do not work: an empirical analysis. *British Journal of Political Science*, 34(04):657–674.

Gneezy, U., Rockenbach, B., and Serra-Garcia, M. (2013). Measuring lying aversion. *Journal of Economic Behavior & Organization*, 93:293–300.

Hart, S. and Kurz, M. (1983). Endogenous formation of coalitions. *Econometrica*, 51(4):1047–1064.

Hume, D. (1752/1994). *Hume: Political Essays*. Cambridge University Press.

Kaufman, S. J., Little, R., and Wohlforth, W. C. (2007). *The balance of power in world history*. Springer.

Kauppi, H. and Widgrén, M. (2004). What determines eu decision making? needs, power or both? *Economic Policy*, 19(39):222–266.

Kubicek, P. (2009). The commonwealth of independent states: an example of failed regionalism? *Review of International Studies*, pages 237–256.

Leech, D. (2002). An empirical comparison of the performance of classical power indices. *Political Studies*, 50(1):1–22.

Magaloni, B. (2008). Credible power-sharing and the longevity of authoritarian rule. *Comparative Political Studies*, 41(4-5):715–741.

Montero, M., Sefton, M., and Zhang, P. (2008). Enlargement and the balance of power: an experimental study. *Social Choice and Welfare*, 30(1):69–87.

Myerson, R. B. (2008). The autocrat's credibility problem and foundations of the constitutional state. *American Political Science Review*, pages 125–139.

Nash, J. F., Nagel, R., Ockenfels, A., and Selten, R. (2012). The agencies method for coalition formation in experimental games. *Proceedings of the National Academy of Sciences*, 109(50):20358–20363.

Owen, G. (1972). Multilinear extensions of games. *Management Science*, 18(5-part-2):64–79.

Ray, D. (2007). *A game-theoretic perspective on coalition formation*. Oxford University Press.

Rode, M., Pitlik, H., and Borrella Mas, M. A. n. (2018). Does fiscal federalism deter or spur secessionist movements? empirical evidence from europe. *Publius: The Journal of Federalism*, 48(2):161–190.

Schelling, T. C. (1960/1980). *The strategy of conflict*. Harvard University Press.

Svolik, M. W. (2009). Power sharing and leadership dynamics in authoritarian regimes. *American Journal of Political Science*, 53(2):477–494.

Tremewan, J. and Vanberg, C. (2016). The dynamics of coalition formation–a multilateral bargaining experiment with free timing of moves. *Journal of Economic Behavior & Organization*, 130:33–46.

Tubilewicz, C. (2016). *Critical Issues in Contemporary China: Unity, Stability and Development*. Taylor & Francis.

Vachudova, M. A. and Hooghe, L. (2009). Postcommunist politics in a magnetic field: How transition and eu accession structure party competition on european integration. *Comparative European Politics*, 7(2):179–212.

van Damme, E. (1998). On the state of the art in game theory: An interview with Robert Aumann. *Games and Economic Behavior*, 24(1):181–210.

Waltz, K. N. (1979/2010). *Theory of international politics.* Waveland Press.

Wolff, S., Ross, S., and Wee, A. (2020). Subnational governance and conflict: The merits of subnational governance as a catalyst for peace.

World Bank (2011). World development report 2011: Conflict, security and development. *World Bank: Washington, DC.*

A Supplementary experimental results

A.1 Overview of the winning coalitions

	Binding agreement		Delayed agreement		Delayed agreement with communication	
	S_7 (7,4,4,4)	S_9 (9,4,4,4)	S_7 (7,4,4,4)	S_9 (9,4,4,4)	S_7 (7,4,4,4)	S_9 (9,4,4,4)
SW	86.4%	94.5%	89.1%	37.0%	90.7%	35.4%
SWW	-	-	4.1%	4.8%	-	1.4%
WWW	13.6%	4.1%	6.8%	50.7%	8.7%	59.9%
SWWW	-	1.4%	-	7.5%	0.6%	3.3%
Total	100%	100%	100%	100%	100%	100%
N coalitions	147	145	147	146	150	147

Table 8: Percentages of winning coalitions by treatment and coalition type.

A.2 Overview of the initial coalition proposals

A.2.1 Proposals made by the weak players

	Binding agreement		Delayed agreement		Delayed agreement with communication	
Weak players	S_7 (7,4,4,4)	S_9 (9,4,4,4)	S_7 (7,4,4,4)	S_9 (9,4,4,4)	S_7 (7,4,4,4)	S_9 (9,4,4,4)
SW	69.8%	72.2%	64.9%	18.4%	69.6%	14.4%
SWW	2.4	3.1	8.9%	13.3%	0.2	5.1%
WWW	23.6%	19.8%	21.6%	54.2%	24.2%	70.7%
SWWW	4.2	4.9%	4.7	14.0%	6.0 %	9.8%

Table 9: Percentages of initial coalitions proposed by weak players, by treatment and coalition type.

A.2.2 Proposals made by the strong players

	Binding agreement		Delayed agreement		Delayed agreement with communication	
Strong players	S_7 (7,4,4,4)	S_9 (9,4,4,4)	S_7 (7,4,4,4)	S_9 (9,4,4,4)	S_7 (7,4,4,4)	S_9 (9,4,4,4)
SW	92.7%	92.7%	74.0%	63.3%	90.7%	56.7%
SWW	-	4.7	18.7%	9.3%	2.0	18.0%
SWWW	7.3	2.7%	7.3	27.3%	7.3%	25.3%

Table 10: Percentages of initial coalitions proposed by strong players, by treatment and coalition type.

A.3 Average payoffs

	Binding agreement		Delayed agreement		Delayed agreement with communication	
	Strong	Weak	Strong	Weak	Strong	Weak
S_7	0.57	0.14	0.48	0.17	0.56	0.15
	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
S_9	0.64	0.12	0.32	0.23	0.25	0.25
	(0.01)	(0.00)	(0.00)	(0.01)	(0.02)	(0.01)
Mean Diff.	0.7	-.02	-0.16	0.06	-0.32	0.10
	(0.02)	(0.01)	(0.03)	(0.01)	(0.03)	(0.01)

Table 11: Average realized payoffs (as proportion of the prize) of a strong and a weak player (standard errors, clustered by group, in brackets).

A.4 Weak players' payoff in SW winning coalitions

	Coef.	Std.Err.
Reference treatment:		
Delayed-S_7 (base)	+0.45***	0.03
Differences (to base):		
Delayed-Comm-S_7	-0.08***	0.02
Delayed-S_9	-0.20***	0.03
Delayed-Comm-S_9	-0.19***	0.03
WWW-Proposed	+0.02	0.02
Weak-Proposed-SW	-0.02	0.02

Note: Estimation by multilevel mixed models, with session and participant random effects. The estimation only uses observations where the winning coalition is SW. The dependent variable is the payoff of the weak player. "WWW-Proposed" is a dummy variable that equals to 1 if the weak player chose to form the SW coalition when both WWW and SW were available, and equals to 0 otherwise. "Weak-Proposed-SW" is a dummy variable that equals to 1 when the SW coalition that formed had been proposed by the weak player, and equals to 0 if the weak player did not propose SW but chose to form the SW coalition. The coefficients are tested against zero and the significance levels are marked as *** 0.1%, ** 1%, * 5%.

Table 12: History effects of proposals on the weak player's payoff in SW coalitions

A.5 Number of proposals needed for coalition formation

A.5.1 Overall distributions

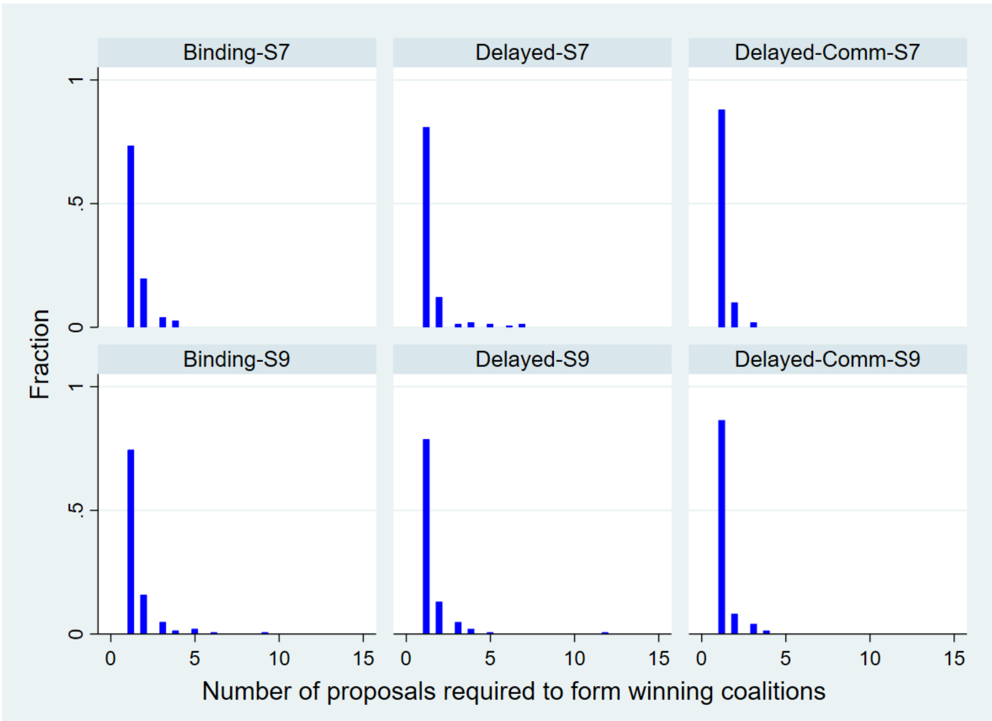

Figure 2: Histograms of the number of proposals required to form a winning coalition

A.5.2 Treatment comparisons and the role of communication

	Coef.	Std.Err.
Reference treatment:		
Binding-S_7 (base)	1.36***	0.07
Differences (to base):		
Binding-S_9	0.04	0.10
Delayed-S_7	-0.01	0.10
Delayed-S_9	-0.12	0.10
Delayed-Comm-S_7	-0.23*	0.10
Delayed-Comm-S_9	-0.22*	0.10

Note: Estimation by mixed model with session random effects. Dependent variable is the number of proposal rounds needed in a given four-player group and period until agreement was reached. The coefficients are tested against zero and the significance levels are marked as *** 0.1%, ** 1%, * 5%.

Table 13: The number of proposals needed to form coalitions across treatments

	Coef.	Std.Err.
Reference treatment:		
Delayed-S_7-SW (base)	1.37***	0.09
Differences (to base):		
Delayed-S_9-SW	-0.03	0.16
Delayed-Comm-S_7-SW	-0.26**	0.13
Delayed-Comm-S_9-SW	-0.29*	0.16
Delayed-S_7-WWW	-0.26	0.26
Delayed-S_9-WWW	-0.20	0.15
Delayed-Comm-S_7-WWW	0.03	0.25
Delayed-Comm-S_9-WWW	-0.18	0.14

Note: Estimation by mixed model with session random effects. Dependent variable is the number of proposal rounds needed in a given four-player group and period until agreement was reached. The coefficients are tested against zero and the significance levels are marked as *** 1%, ** 5%, * 10%.

Table 14: The number of proposals needed to form SW and WWW coalitions in *Delayed*

A.6 Messages sent in treatments with communication

	No-message	Message w/o split proposal	Message with split proposal	Select other's proposal	Total
Delayed-with-comm-S_7	66	50	188	10	314
Delayed-with-comm-S_9	99	73	202	20	394
Total	165	123	390	30	718

Note: "Select other's proposal" refers to the situation when a player is part of the winning coalition but the coalition formed is not the same as the coalition proposed by this player. In this case, the split proposal that may be contained in the message sent by this player together with her own coalition proposal would not match with the actual split proposal made by her after the coalition is formed. Hence it is not possible to check if this player was trustworthy or not.

Table 15: Type of messages sent with the coalition proposals in the round when a winning coalition is formed

A.7 Weak players' willingness to include the strong player, controlling for risk attitudes

	Weak players' likelihood to include the strong player							
	Model 1		Model 2		Model 3		Model 4	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Binding-S_7 (Constant)	+0.77***	0.07	+0.70***	0.15				
Binding-S_9	+0.04	0.10	+0.05	0.10				
Delayed-S_7	+0.03	0.10	+0.03	0.10	+0.79***	0.08	+0.61***	0.19
Delayed-S_9	-0.31**	0.10	-0.31**	0.10	-0.33**	0.12	-0.34**	0.12
Delayed-Comm-S_7	-0.01	0.10	-0.01	0.10	-0.04	0.12	-0.04	0.12
Delayed-Comm-S_9	-0.47***	0.10	-0.46***	0.10	-0.50***	0.12	-0.48***	0.12
Risk-taking			+0.01	0.01			+0.01	0.01
Male			-0.08*	0.03			-0.04	0.04
Other Controls	No		Yes		No		Yes	

Note: Estimation by multilevel mixed models, with session and participant random effects. Dependent variable is a dummy that equals 1 if a weak player's initial coalition proposal in a given round included the strong player, and equals 0 otherwise. In models 1 and 2, *Binding-S7* is the reference group, whereas in models 3 and 4 *Delayed-S7* is the baseline group. "Other Controls" include age and measures of altruism and trust. The coefficients are tested against zero and the significance levels are marked as *** 0.1%, ** 1%, * 5%.

Table 16: Weak players' likelihood to include the strong player in their initial proposals. We control for risk attitudes using our measure of risk preferences

To assess whether our results could be driven by risk attitudes, we elicited risk preferences, following Dohmen et al. (2011)'s approach. We asked participants to answer the question "How willing are you to take risks, in general?", using a Likert scale taking values for 1 to 10. Table 16 shows the effects of different treatments when controlling for the risk attitudes of participants. All our results are left unchanged.

B A simple model of ex ante communication

Consider the variant with delayed agreement where the players can, together with their coalition proposal, send a message to the players who are part of the proposal. Sending a message has no direct cost. We assume that the players' messages are split proposals $\boldsymbol{\mu}$. In the minimal winning coalition (W, W, W), an equal split is the natural outcome to expect, with and without communication. Thus, the analysis below focuses on the effect of ex ante communication on the emergence (S, W) coalitions.

Suppose that some share $\alpha \in [0,1]$ of players have a non-monetary cost of deviating from their proposed split; for simplicity, we call this cost a "cost of lying" in what follows and assume additive separability of the utility function. If μ_i is the prize share that player i claimed for herself in her message and β_i is the actual share obtained, the cost of lying is given by

$$c(\mu_i, \beta_i) = \frac{1}{2} c_L \left(\max\{\beta_i - \mu_i, 0\}\right)^2,$$

that is, i experiences a positive cost of lying if $c_L > 0$ and $\mu_i < \beta_i$ (she claimed a smaller prize share ex ante than she finally demanded/obtained). Note that for simplicity we assume the cost of lying to be independent of the process of how the final split $\boldsymbol{\beta}$ has been decided.[43]

Communication in the *Delayed-Comm-*S_9 treatment. Consider first the case of $2m_W < m_S$. There is always a "babbling" equilibrium in which the communication is ignored. To derive testable predictions on a possible effect of communication, suppose the strong player proposes (S, W_k) together with a message μ_S, independent of her 'type' (cost of lying). If weak player k accepts, she expects a payoff of

$$(1-\alpha)\, 0 + \alpha \left(1 - \beta_S^{\alpha}\right)$$

where β_S^{α} is the prize share that a strong player claims in stage 2 in case she has cost of lying $c_L > 0$. (Strong players with zero cost of lying will claim a share $\beta_S = 1$, which the weak

[43] Assuming that the cost of lying is increasing in the difference between announced intention and actual demand when bargaining and that the prize share obtained is increasing in the actual demand would yield similar conclusions.

player rationally expects.) Given the message μ_S, β_S^α is the solution to

$$\max_{\beta_S} \beta_S - \frac{1}{2} c_L \left(\max\left\{\beta_S - \mu_S, 0\right\}\right)^2 \text{ s.t. } \beta_S \in [0,1],$$

which yields

$$\beta_S^\alpha = \min\left\{\frac{1 + c_L \mu_S}{c_L}, 1\right\}. \tag{4}$$

The amount β_S^α which the strong player claims in stage 2 is increasing in her announcement μ_S (decreasing in the share $1 - \mu_S$ she promised to the weak player) and decreasing in the cost of lying c_L.[44] If $c_L \to \infty$ then $\beta_S^\alpha \to \mu_S$: for infinitely high cost of lying, the strong player sticks to her promise.

Anticipating the stage 2 choice of the types of strong players with and without cost of lying, the weak player accepts the strong player's proposal if and only if it is (weakly) preferred to the alternative coalitions that have been proposed. Suppose that (W, W, W) has been proposed and that weak player k expects a prize share of $1/3$ in the (W, W, W) coalition. Then, with (4), weak player k prefers the proposal (S, W_k) if and only if

$$\alpha \left(1 - \min\left\{\frac{1 + c_L \mu_S}{c_L}, 1\right\}\right) \geq \frac{1}{3}.$$

Necessary conditions for this inequality to hold are $\alpha \geq 1/3$ and $(1 + c_L \mu_S)/c_L < 1$, that is, $\mu_S < (c_L - 1)/c_L$. In the latter case, the condition for acceptance of the proposal is equivalent to

$$\mu_S \leq \frac{3\alpha - 1}{3\alpha} - \frac{1}{c_L}. \tag{5}$$

Since μ_S must be non-negative, a pooling equilibrium with message

$$\mu_S^* = \frac{3\alpha - 1}{3\alpha} - \frac{1}{c_L} \tag{6}$$

in which a (S, W) coalition is formed exists only if $\alpha > 1/3$ and c_L is sufficiently large.[45] If

[44] In (S, W) coalitions of stage 2, a payoff-maximising player W accepts all splits $1 - \beta_S$ since, the strong player S, due to her super-majority, can keep the entire prize for herself if she wants.

[45] If $\alpha > 1/3$, $\mu_S^* \geq 0$ is equivalent to $c_L \geq 3\alpha/(3\alpha - 1)$. Note that condition (5) implies $\mu_S < (c_L - 1)/c_L$. A complete equilibrium characterisation requires assumptions on off-equilibrium beliefs. We can assume, for instance, that the weak players believe that the strong player has zero cost of lying in case $\mu_S < \mu_S^*$ or no

$c_L \to \infty$ (i.e., a share α of strong players does not deviate from the initial announcement) and $\alpha \to 1$, the strong player needs to promise a share of at least $1 - \mu_S = 1/3$ to the weak player in order to induce her to accept the (S, W) coalition, exactly what W could get in the (W, W, W) coalition. If $\alpha < 1$, the required promise $1 - \mu_S$ to the weak player goes up since W anticipates that some strong players will not stick to their promise. For instance, if $\alpha = 2/3$ and $c_L \to \infty$, the strong player needs to promise an equal split in order for a (S, W) coalition to form. Upon acceptance, W indeed gets a prize share of $1/2$ with probability $\alpha = 2/3$ and gets zero with the remaining probability. This makes her exactly indifferent to what she can expect in a (W, W, W) coalition. We summarise these considerations as follows:

Observation 1 *Let $2m_W < m_S$ as in the S_9 treatments.*
(i) If $\alpha > 1/3$ and $c_L \geq 3\alpha/(3\alpha - 1)$, there is a (pooling) equilibrium in which the strong player proposes a coalition (S, W_k) together with a message given by (6) and the coalition (S, W_k) forms, with an expected payoff of weak player k equal to $1/3$.
(ii) If $\alpha \leq 1/3$ or $c_L < 3\alpha/(3\alpha - 1)$, all weak players j strictly prefer a (W, W, W) coalition over a coalition (S, W_j), independent of the strong player's message μ_S.

Accordingly, a sufficiently high share of players with a sufficiently high non-monetary cost of deviating from the initial announcement (as in part (i) of Observation 1) can constitute an explanation for why the likelihood of (S, W) coalitions may be higher in the *Delayed-Comm-S_9* treatment than in the *Delayed-S_9* treatment (Hypothesis 3B(i)). However, if α and/or c_L is low (as in part (ii) of Observation 1), ex ante communication may be ineffective (Hypothesis 3A).

Communication in the *Delayed-Comm-*S_7 treatment. Suppose that $2m_W \geq m_S$ so that with delayed agreement the Shapley value in (S, W) coalitions is $(1/2, 1/2)$. For the weak players, (S, W) coalitions hence yield the highest expected payoff. In fact, if weak players expect a payoff of $1/3$ under an equal split in a (W, W, W) coalition, weak player k may even be willing to accept less than $\beta_k = 1/2$ in a coalition (S, W_k), which she may attempt to indicate through her ex ante announcement in order to be selected by the strong player with higher probability.

message is sent, and has positive costs of lying $c_L > 0$ with probability α in case $\mu_S > \mu_S^*$.

To derive again a testable prediction, consider a weak player k who proposes a coalition (S, W_k) together with a message μ_k, where μ_k denotes the prize share which weak player k claims for herself. In contrast to the previous case where the strong player alone had a super-majority in (S, W) coalitions, we need an assumption on how the final allocation $\boldsymbol{\beta}$ depends on the players' demands, which, in turn, may depend on the messages. To obtain closed-form solutions, we simply assume that weak player k's prize share β_k in a coalition (S, W_k) is equal to her demand if the demand is below 1/2 (that is, if W_k makes concessions) and is equal to 1/2 otherwise. This means that a weak player with zero cost of lying would obtain a prize share of 1/2 in a (S, W) coalition, just as in the case without communication. A weak player k with a positive cost of lying who sent a split proposal μ_k (indicating her own prize share) would–in a winning coalition (S, W_k)–demand a quantity that is the solution to

$$\max_{\beta_k} \beta_k - \frac{1}{2} c_L \left(\max\left\{\beta_k - \mu_k, 0\right\}\right)^2 \text{ s.t. } \beta_k \in \left[0, \frac{1}{2}\right].$$

This yields

$$\beta_k^{\alpha} = \min\left\{\frac{1 + c_L \mu_k}{c_L}, \frac{1}{2}\right\} \tag{7}$$

as the prize share demanded (and obtained) by a weak player k who sent a message μ_k and has a positive cost of lying. In a pooling equilibrium where all weak players j send the same message μ_j together with a coalition proposal (S, W_j), the strong player expects a prize share $\beta_S = (1 - \alpha)(1/2) + \alpha(1 - \beta_k^{\alpha})$ in coalition (S, W_k), which is increasing in the share $1 - \mu_k$ that weak player k promised to the strong player. This holds as soon as $\alpha > 0$ (possibly small) even though the strong player correctly anticipates that the weak player may deviate from her initial announcement. Since both players in (S, W) coalitions have equal bargaining power, accepting a coalition offer (S, W_k) with a high promise $1 - \mu_k$ is attractive for the strong player even when α is low: under the above assumption on the stage 2 bargaining, the worst that can happen to the strong player is that the coalition reverts to a split $(1/2, 1/2)$. For the strong player's expected benefit to be strictly larger than 1/2 (that is, for $\beta_k^{\alpha} < 1/2$), c_L must be sufficiently large (to be precise, $c_L > 2$). Otherwise, the weak player k would request an equal split at stage 2 even for the lowest possible message $\mu_k = 0$.

Multiple equilibria can be supported by appropriate off-equilibrium beliefs.[46] Two considerations may favour a particular type of equilibrium. First, there may be "price competition" among the weak players in the sense that they lower the payoff claim in their message. Such competitive effects emerge if the strong player selects coalition offers accompanied by more favourable promises with higher probability.[47] Second, in an equilibrium with message μ_k^*, a weak player with a cost of lying should (at least) get a payoff of $\beta_k^\alpha = 1/3$ (where β_k^α is given in (7)), which is what she could expect in a (W, W, W) coalition. This, together with feasibility constraint $\mu_k \geq 0$, characterises the most generous promise

$$\mu_k = \max\left\{\frac{c_L/3 - 1}{c_L}, 0\right\} \tag{8}$$

to the strong player, which is still credible in the sense that weak players with a cost of lying (weakly) prefer a coalition (S, W) over a coalition (W, W, W). This possibility of competition for (S, W) coalitions among the weak players can be summarised as follows:

Observation 2 *Let $2m_W \geq m_S$ as in the S_7 treatments. If $\alpha > 0$ and $c_L > 2$, there is a (pooling) equilibrium in which all weak players propose a coalition (S, W) together with a message given by (8) and the strong player's expected payoff is larger than* $1/2$.

In equilibrium, the weak players may announce that they intend to demand a share smaller than $1/2$. If some players have a sufficiently high non-monetary cost of deviating from the initial announcement, ex ante communication increases the strong player's expected payoff beyond her Shapley value of $1/2$ (Observation 2). If c_L is low, however, even those players with a (small) cost of lying find it too attractive to deviate from their promise so that the strong player does not get more than $1/2$ in the equilibrium with ex ante communication. In contrast to the requirements on the role of communication in case of $2m_W < m_S$ (Observation 1), the share α of players with a cost of lying does not need to be high for communication to be effective. The strong player's expected payoff, however, is increasing in α in the equilibrium characterised in Observation 2. Also, even when competition

[46] To illustrate, there is a set of pooling equilibria in which all weak players j send a message $\mu_j^* < 1/2$ together with their coalition proposal (S, W_j), supported by sufficiently pessimistic beliefs of the strong player who, upon observing a message $\hat{\mu}_k \neq \mu_j^*$, believes that k's cost of lying is zero.

[47] Deviations to lower messages $\hat{\mu}_k$ are profitable, for instance, if the strong player's posterior belief still assigns probability α to the case that weak player k has a cost of lying $c_L > 0$.

between the weak players is less fierce or there is a lower bound on the messages used by the weak players j (e.g., $\mu_j \geq 1/3$), the strong player's payoff is strictly larger than $1/2$ if c_L is sufficiently large (compare (7)). Thus, some players with a (non-monetary) cost of deviating from the initial announcement can constitute an explanation for why the strong player's average realised payoff may be higher in the *Delayed-Comm-*S_7 treatment than in the *Delayed-*S_7 treatment (Hypothesis 3B(ii)).

C Experimental instructions and screenshots

C.1 Instructions for the *Binding*-S_9 treatment

Welcome to our experiment and thank you for taking part. In this experiment, in addition to your show up fee (RMB 20), it is possible to earn additional money depending on both your choices and the choices of other participants in this room. It is important that you read these instructions carefully. If you have any questions please raise your handan experimenter will come to answer your question at your desk. During this experiment, please switch off your mobile phone, and please do not communicate with any other participants unless the experiment specifically allows it. All information collected during this experiment is confidential and anonymous.

Your task

In the experiment, you will play 15 rounds of a task described by steps 1 to 4 below. At the beginning of each round, the computer will randomly allocate you into a group of four participants in the room, and a player number between 1 to 4 will be randomly assigned to you. In each four-player group: there are 21 votes in total; three participants will have 4 votes each and one participant will have 9 votes; and the computer will randomly determine whether you will have 4 votes or 9 votes.

Your task in each round is to propose and decide on how to allocate 48 Experimental Currency Units (ECU) among the participants in your own group. This allocation process takes place as follows.

1. You will need to select a set of player(s) (including yourself) and specify an amount you would like to allocate to each player you select. The set of players you propose can contain any number of players but must have more than 1/2 of the total votes of the four-player group (i.e., at least 11 votes).

2. Once all participants have made their proposals, each player will see all the proposals that include him/her.

3. Each player will then choose his/her most preferred proposal. A proposal will only be agreed/passed if all players who are part of that proposal have chosen it.

4. If none of the proposals is agreed, there is a 5% chance that the round will end and

all players in the group will receive zero payment. There is a 95% chance that all players in the group can make new proposals (starting from step 1 above) and try to agree on an allocation again.

Before we start this experiment, you will need to answer some control questions to ensure that you have fully understood the instructions. Afterwards, you will have the opportunity to do one practice round to familiarize yourself with the task and the computer screens you will see in the experiment. Again, at the beginning of each round, your group, your player number and the number of votes you have will be randomly re-assigned. At the end of each round, you will be informed the final decision outcome for your own group irrespective whether you are part of the agreed proposal/allocation.

After you finish all 15 rounds, the computer will select one out of the 15 rounds at random; Your earnings from this round will be converted to RMB (at a rate of 1 ECU = 2 RMB), which will be paid to you (together with your show up fee (20 RMB).

C.2 Experimental screenshots for the *Binding*-S_9 treatment

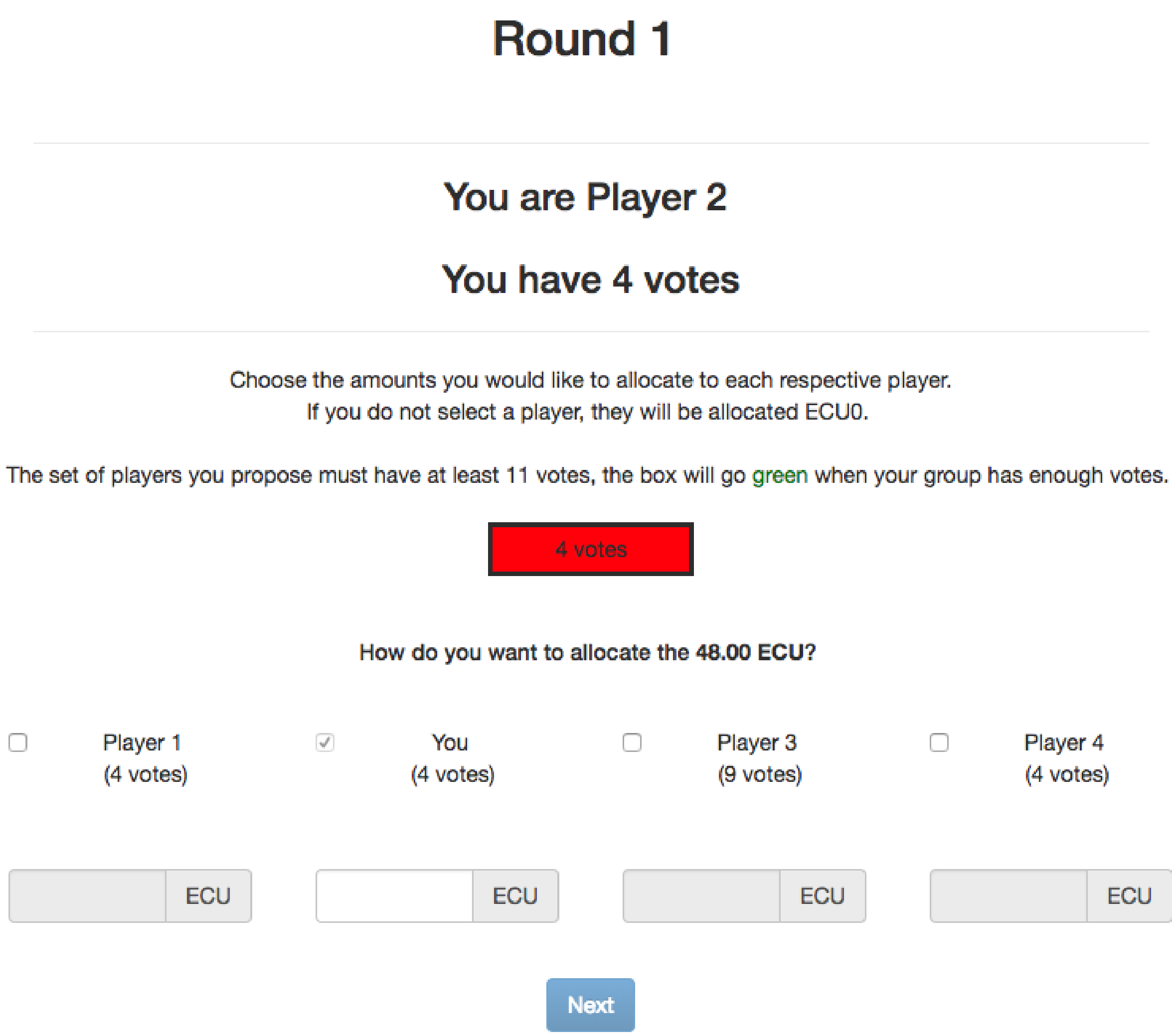

Figure 3: A player proposes a coalition with a split in the *Binding*-S_9 treatment.

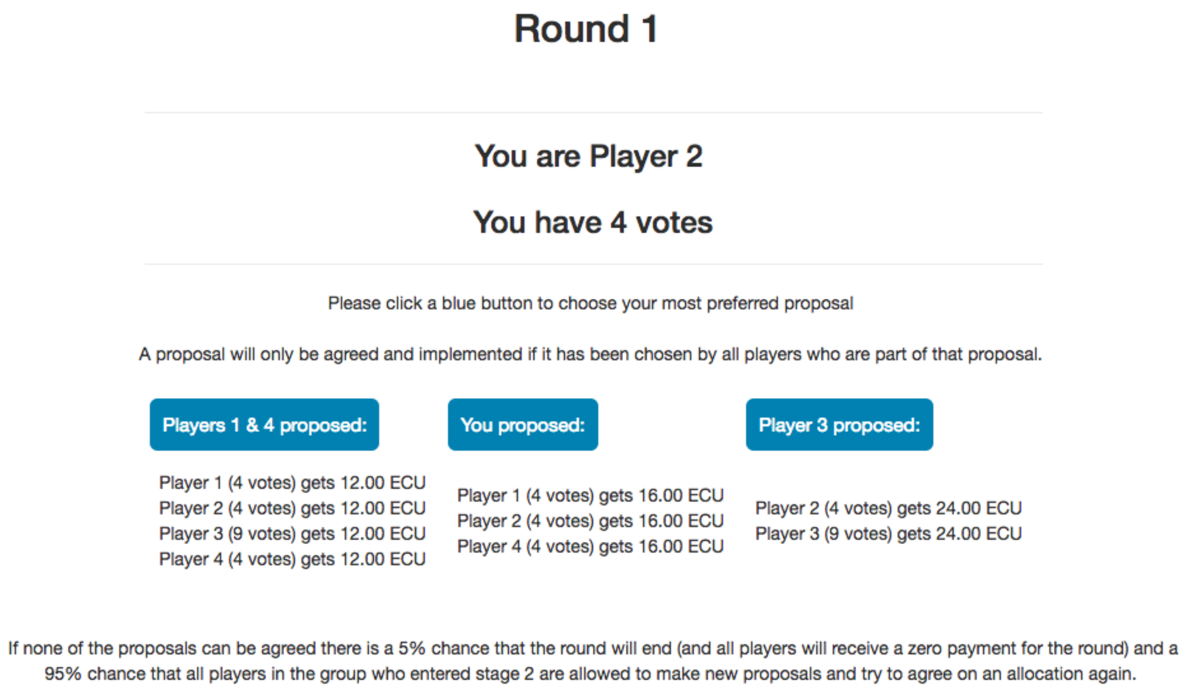

Figure 4: A player selects the most preferred proposal from all proposals that include herself in the *Binding-*S_9 treatment.

C.3 Instructions for the *Delayed-Comm-*S_9 treatment

Welcome to our experiment and thank you for taking part. In this experiment, in addition to your show up fee (RMB 20), it is possible to earn additional money depending on both your choices and the choices of other participants in this room. It is important that you read these instructions carefully. If you have any questions please raise your handan experimenter will come to answer your question at your desk. During this experiment, please switch off your mobile phone, and please do not communicate with any other participants unless the experiment specifically allows it. All information collected during this experiment is confidential and anonymous.

Your task

In the experiment, you will play 15 rounds of a task described by steps 1 to 7 below. At the beginning of each round, the computer will randomly allocate you into a group of four participants in the room, and a player number between 1 to 4 will be randomly assigned to you. In each four-player group: there are 21 votes in total; three participants will have 4 votes each and one participant will have 9 votes; and the computer will randomly determine whether you will have 4 votes or 9 votes.

Your task in each round is to propose and decide on how to allocate 48 Experimental Currency Units (ECU) among the participants in your own group. This allocation process takes place in two stages. In stage 1, you will propose and decide who could enter stage 2, whereas in stage 2, you will propose and decide how to allocate 48 ECU among those who have entered stage 2.

1. In the first stage, you need to select a set of players (including yourself) who could enter stage 2. The set of players you propose can contain any number of players but must have more than 1/2 of the total votes of the four-player group (i.e., at least 11 votes). *Together with your proposal, you have the option to send a text message to the players you selected. The content of your message is not restricted in any way, except that you are forbidden to reveal your identity, seat number or anything that might uncover your anonymity, or to make threats. If you violate these restrictions you will be asked to leave the laboratory and will not be paid.*

2. Once all participants have made their proposals, each player will see all the proposals *(together with any text messages)* that include him/her.[48]

3. Each player will then choose his/her most preferred proposal. A proposal will only be agreed if all players who are part of that proposal have chosen it. All participants in the agreed proposal will enter stage 2. If you are not in the agreed proposal, then you will take no further part in the round and you will receive zero payment for that round.

4. If none of the proposals is agreed, there is a 5% chance that the round will end and all players in the group will receive zero payment. There is a 95% chance that all players in the group can make new proposals (starting from step 1 above) and try to agree on who could enter stage 2.

(The following steps only apply to those who will enter stage 2.)

5. In the second stage, you need to propose and decide the allocation of the 48 ECU only among those who have entered stage 2. You can select any number of player(s) and

[48] The instructions for the *Delayed-*S_9 treatment is the same except the *italizised* part of the information regarding communication in step 1 and 2 are removed.

specify an amount you would like to allocate to each player you select. When selecting other players, you need to make sure that you - together with the other players you might have selected - have more than 2/3 of the total votes of all players who entered stage 2.

6. Once all players have made their proposal, each player will see all proposed allocations that include him/her. Each player will then choose his/her most preferred allocation. An allocation will only be agreed and implemented if all players who are part of it have chosen it.

7. If none of the proposed allocation is agreed, there is again a 5% chance that the round will end and all players in the group will receive zero payment. There is a 95% chance that all players who entered stage 2 can make new proposals starting from step 5 above and try to agree on an allocation again.

Before we start this experiment, you will need to answer some control questions to ensure that you have fully understood the instructions. Afterwards, you will have the opportunity to do one practice round to familiarise yourself with the task and the computer screens you will see in the experiment. Again, at the beginning of each round, your group, your player number and the number of votes you have will be randomly re-assigned. At every stage of the game, you will be informed the final decision outcome for your own group irrespective whether you are part of the agreed proposal/allocation.

After you finish all 15 rounds, the computer will select one out of the 15 rounds at random; Your earnings from this round will be converted to RMB (at a rate of 1 ECU = 2 RMB), which will be paid to you (together with your show up fee (20 RMB).

C.4 Experimental screenshots for the *Delayed-Comm-*S_9 treatment

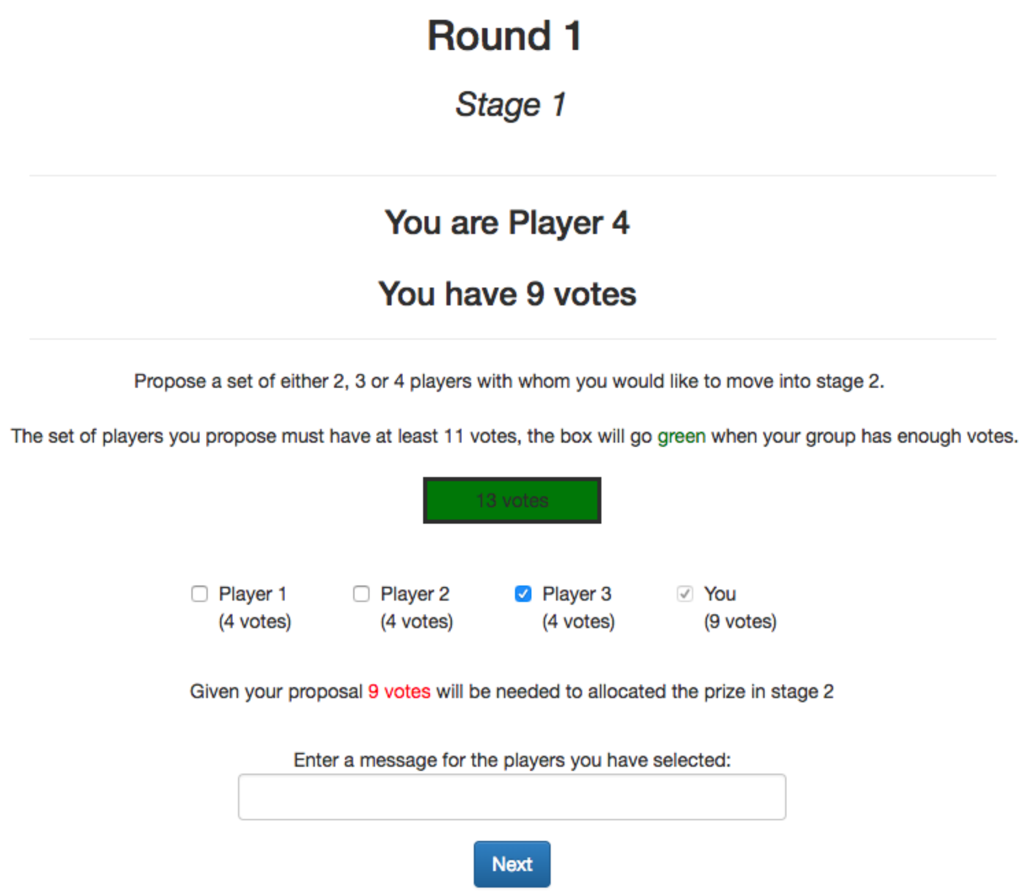

Figure 5: A player proposes a coalition to be formed in stage 1 in the *Delayed-Comm-*S_9 treatment.

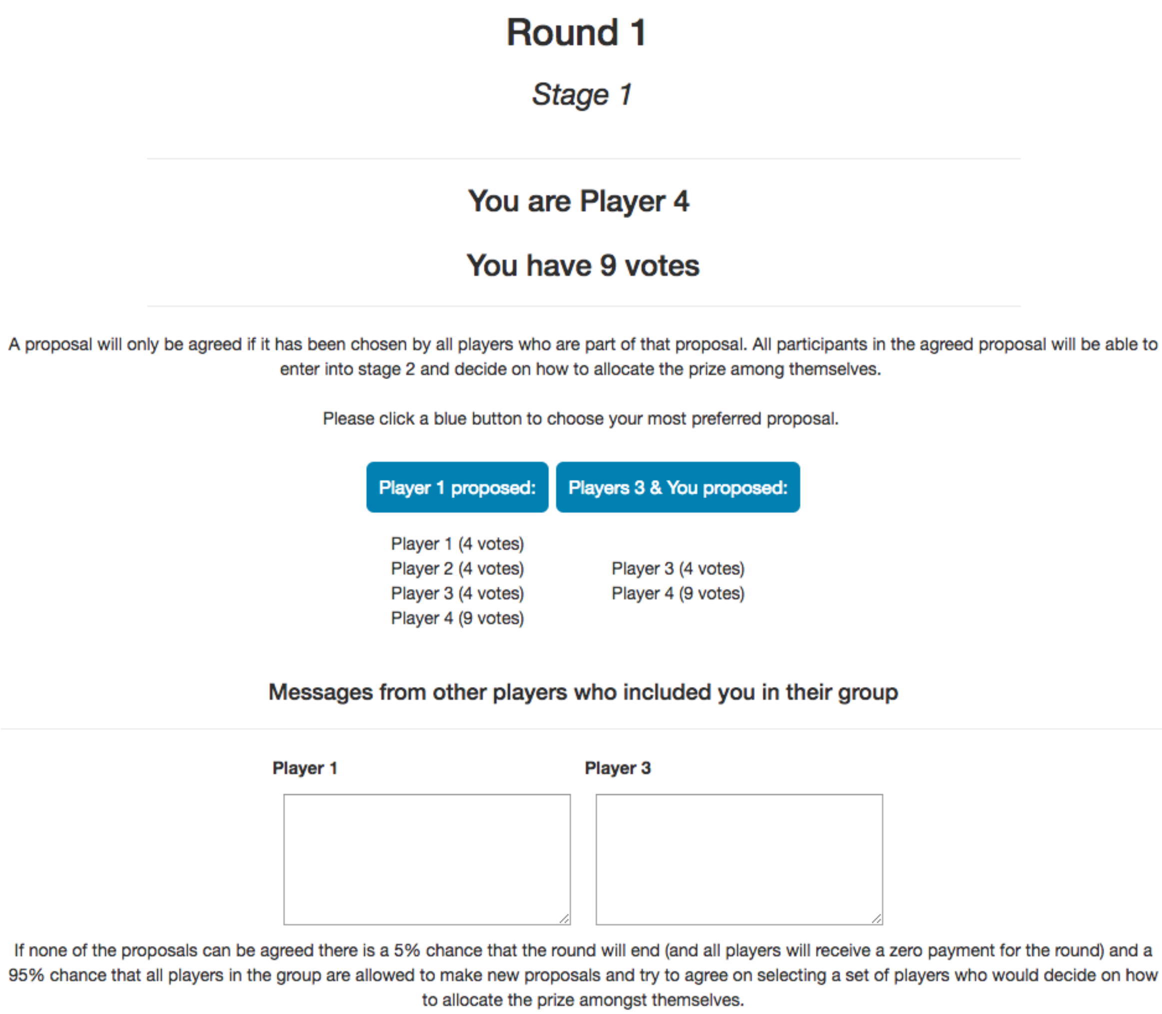

Figure 6: A player selects a preferred coalition that include herself in stage 1 in the *Delayed-Comm-*S_9 treatment.

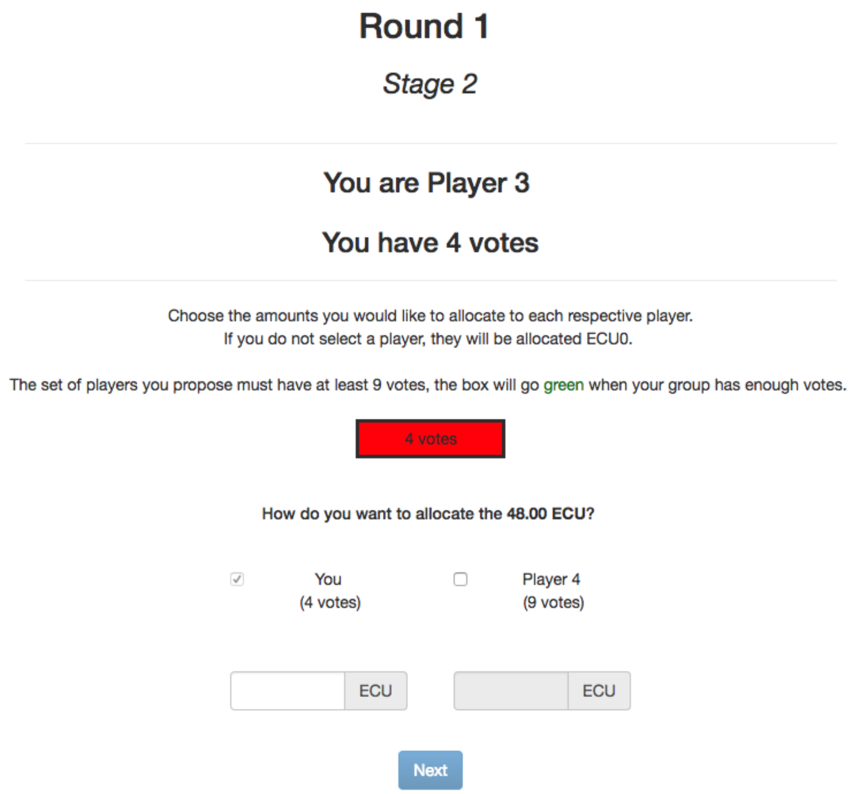

Figure 7: A player in the winning coalition proposes a split in stage 2 in the *Delayed-Comm-*S_9 treatment.

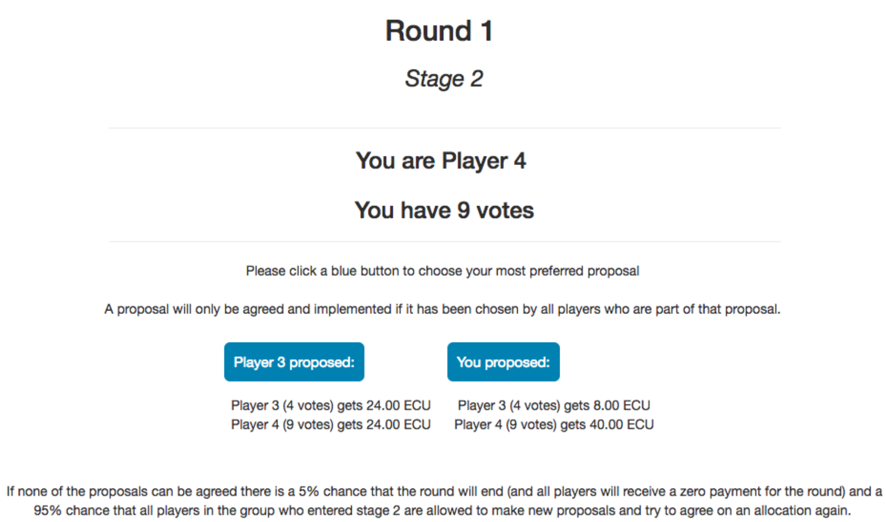

Figure 8: A player in the winning coalition selects a split that includes herself in stage 2 in the *Delayed-Comm-*S_9 treatment.