Online Appendices

A Derivation of behavioral hypotheses

Expected allocation after coalitional commitment

We use a backward induction argument, starting with the game after a coalitional commitment has occurred. Following a coalitional commitment to $W \in \mathcal{W}$, the predicted outcome is the proportional allocation $x^p(W)$.

The Nash bargaining solution (NBS) is the natural cooperative game-theoretic counterpart of the stable set when considering pure (within-coalition) bargaining rather than coalitional (between-coalition) bargaining. The NBS predicts proportionality/equality in our setting without outside options. The NBS for n players is characterized by the maximization problem $\max_x \prod_{i \in W} u_i(x)$ subject to $\sum_{i \in N} x_i \leq 100$ (Nash, 1950; Harsanyi and Selten, 1972; Okada, 2010). Suppose by contradiction that $u_i(x) \neq u_j(x)$ for some representatives $i, j \in W$ and let $\bar{u}$ be the mean of $u_i(x)$ and $u_j(x)$. Thus, $u_i(x) = \bar{u} + d$ and $u_j(x) = \bar{u} - d$ for some $d \neq 0$. In addition, $u_i(x) + u_j(x) = 2\bar{u}$ and $u_i(x)u_j(x) = \bar{u}^2 - 2d < \bar{u}^2$. Replacing both $u_i(x)$ and $u_j(x)$ by $\bar{u}$ increases the product of payoffs while keeping the sum of payoffs fixed. The NBS is therefore achieved when $d = 0$ and $u_i(x) = u_j(x)$ for all $i, j \in W$ and $u_i(x) = 0$ for all $i \notin W$—the proportional solution.

Expected allocations in stage 2

We next consider stage 2 of the coalitional bargaining game. All allocations can be reached via an allocative commitment. Coalitional commitment thus plays no role in stage 2 from a theoretical perspective.

The unique stable set in the coalitional weighted majority game is the main simple solution, X^a (e.g., von Neumann and Morgenstern, 1944; Ray and Vohra, 2015a). One can verify internal and external stability of $\{(50, 50, 0), (50, 0, 50), (0, 50, 50)\}$ in the three-party setting and of $\{(33^1/_3, 0, 0, 66^2/_3), (0, 33^1/_3, 0, 66^2/_3), (0, 0, 33^1/_3, 66^2/_3), (33^1/_3, 33^1/_3, 33^1/_3, 0)\}$ in the four-party setting. There are typically also discriminatory stable sets in addition to the main simple solution in weighted majority games (e.g., Ray and Vohra, 2015b). However, discriminatory stable sets disappear with discrete allocations. We avoid a proof for brevity.

We thus predict an allocation belonging to the main simple solution, $x \in X^a$, when representatives negotiate in stage 2. Pivotality takes precedence. Proportional allocations, $x \in X^*$, could still be expected in stage 2 if $X^a = X^*$. However, $X^a = X^*$ occurs only for the particular case when vote shares exactly correspond to the so-called homogenous representation of the game: when all MWCs have the same sum of vote shares or, equivalently, when all MWCs are also LWCs (e.g., Morelli and Montero, 2003; Montero, 2017; Eraslan and Evdokimov, 2019).

Prediction 1: The main simple solution determines allocations in stage 2 of all treatments.

Expected allocations in stage 1

In stage 1 of the *Stages1&2* treatments, representatives can commit to coalitions. We showed that a coalitional commitment to winning alliance W leads to the proportional allocation $x^p(W)$. In addition, representatives may choose to forgo coalitional commitment in stage 1 to enter stage 2. In stage 1, they thus consider the *expected* stage-2 allocation. Following Prediction 1, expected stage-2 allocations correspond to the expected main simple solution. Therefore, recalling that a_i is constant across MWCs, we can denote the expected stage-2 allocation by $x^e \equiv (\mu_1 a_1, ..., \mu_n a_n) \in X$, where $\mu_i \in (0, 1)$ is party i's belief that she will be part of the winning coalition in stage 2. The set of relevant allocations in stage 1 is thus $X^p \cup x^e$.

Is there a stable set in stage 1? We show that X^* is the unique candidate for a stable set $Z \subseteq X^p$. That is, only LWCs can be part of a proportional stable set. To see this, note that $x_i^p(W) = v_i/v_W$ for all $i \in W$ implies $x_i^p(W) > x_i^p(W')$ for all $i \in W$ and $W \in \mathcal{W}^*$, $W' \notin \mathcal{W}^*$ (because v_W is the smallest for LWCs). Thus, $x^p(W)$ for $W \in \mathcal{W}^*$ cannot be dominated by any allocation in X^p. By external stability, $x^p(W)$ for $W \in \mathcal{W}^*$ must be part of any stable set $Z \subseteq X^p$. In addition, $x^p(W')$ for $W' \notin \mathcal{W}^*$ is dominated by any $x^p(W)$ with $W \in \mathcal{W}^*$. By internal stability, $x^p(W')$ for $W' \notin \mathcal{W}^*$ cannot be in a stable set that includes some $x^p(W)$, $W \in \mathcal{W}^*$. It follows that X^* is the unique candidate for a stable set $Z \subseteq X^p$.

To constitute a stable set in stage 1, X^* also needs to be externally stable, which means it must dominate allocation x^e (the expected stage-2 outcome that can be used to block allocations in stage 1). By definition, $x^p(W)$ dominates x^e if $x_i^p(W) > x_i^e \Leftrightarrow v_i/v_W > a_i\mu_i$ for all $i \in W$ for some $W \in \mathcal{W}^*$. We verify this requirement for our experimental games. To do so, we must consider specific beliefs μ. The most natural beliefs are $\mu_i = m_i/m$, where m is the total number of MWCs and m_i is the number of MWCs that include party i, reflecting that the main simple solution does not discriminate between different MWCs; each one is equally likely to occur.

Consider the three-party negotiation environment. The set of MWCs consists of coalitions $\{1, 2\}$, $\{1, 3\}$ and $\{2, 3\}$. Only the first two MWCs are LWCs. The proportional allocations are $x^p(\{1, 2\}) = (33, 67, 0)$, $x^p(\{1, 3\}) = (33, 0, 67)$ and $x^p(\{2, 3\}) = (0, 50, 50)$. The main simple solution allocates 50 to each party in a MWC. The expected stage-2 allocation is $x^e = (33^1\!/\!_3, 33^1\!/\!_3, 33^1\!/\!_3)$ because $m = 3$ and $m_i = 2$ such that $\mu_i a_i = 2/3 * 50 = 1/3$ for all i. The MWC-allocation $x^p(\{2, 3\})$ would dominate x^e but is excluded by internal stability. Strictly speaking, neither $x^p(\{1, 2\})$ nor $x^p(\{1, 3\})$ dominate x^e because the lowest proportional payoff is exactly the same as the expected stage-2 payoff. So, this is a knife-edge case. However, a small degree of risk aversion would imply an expected stage-2 utility of less than 1/3 such that x^e would be dominated by both $x^p(\{1,2\})$ and $x^p(\{1,3\})$. The set of LWC allocations is then externally stable in stage 1 and coalitional commitment is expected to occur.

Consider now the four-party negotiation environment. The set of MWCs consists of coalitions $\{1,4\}$, $\{2,4\}$, $\{1,2,3\}$ and $\{3,4\}$. The first three MWCs are LWCs. The proportional allocations are $x^p(\{1,4\}) = (25,0,0,75)$, $x^p(\{2,4\}) = (0,25,0,75)$, $x^p(\{1,2,3\}) = (25,25,50,0)$ and $x^p(\{3,4\}) = (0,0,40,60)$. The expected main simple solution allocates 66²⁄₃ to the large party and 33¹⁄₃ to the other parties in an MWC. The expected stage-2 allocation is $x^e = (16^2/_3, 16^2/_3, 16^2/_3, 50)$, because $m = 4$, $m_i = 2$ for $i = 1,2,3$, and $m_4 = 3$ such that $\mu_i a_i = 1/2 * 33^1/_3 = 16^1/_3$ for $i = 1,2,3$ and $\mu_4 a_4 = 3/4 * 66^2/_3 = 50$. One can see that $x^p(\{1,4\})$ and $x^p(\{2,4\})$ dominate x^e. Together with the third LWC-allocation $x^p(\{1,2,3\})$, they constitute a stable set in stage 1. We thus expect coalitional commitment in stage 1 to occur because it is part of a stable outcome.

In fact, this analysis also implies that there cannot be a stable set in stage 1 that includes x^e. The set X^* is thus the unique stable set in stage 1.

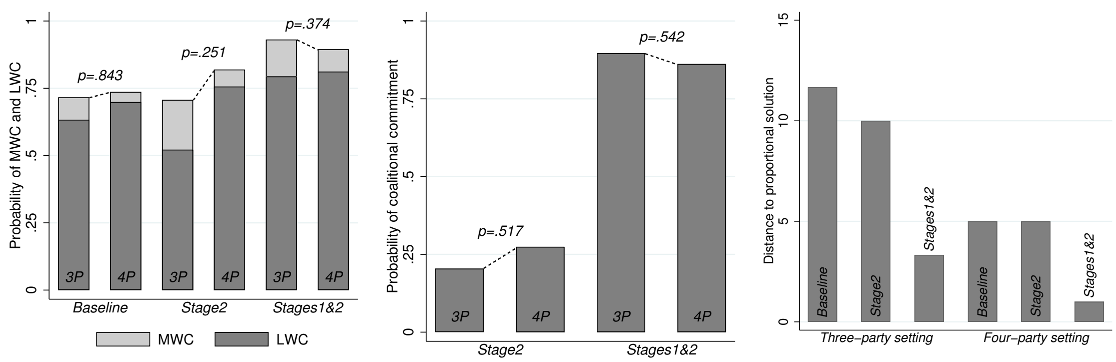
Prediction 2: Coalitional commitment occurs in stage 1 of the *Stage1&2* treatments. The winning coalition is predicted to be an LWC, and its members share the pie proportionally.

B Robustness checks

B.1 Three-Party and Four-Party Setting

Figure 5: Three-party and four-party setting

(a) MWC and LWC (b) Coalitional Commitments (c) Pie Allocations



Notes: ***Figure (a)*** shows the probability of observing a MWC/LWC. ***Figure (b)*** shows the probability of observing coalitional commitments. ***Figure (c)*** shows the median difference between the empirical pie shares and the proportional solution. All P-values are from logit random effects regressions with standard errors clustered on matching groups.

In the main analysis, we pool the data from the three-party and four-party treatments. This approach is justified because our hypotheses equally apply to both settings. Examining if Results 1 to 3 hold independently of the number of parties serves as a helpful robustness check.

Figure 5a shows the probability of observing MWCs and LWCs. As can be seen, there are no significant differences depending on the number of parties. Treatments *3P–Stages1&2* and *4P–Stages1&2* have the highest rates of MWCs and LWCs, but negotiations tend to lead to MWCs and LWCs in all treatments.

Figure 5b shows the probability of observing coalitional commitments. As can be seen, coalitional commitments are common in *3P–Stages1&2* and *4P–Stages1&2* and infrequent in *3P-Stage2* and *4P-Stage2*. Again, there are no significant differences between the three-party and four-party treatments.

Finally, Figure 5c shows the median distance between the empirical pie shares and the proportional solution. The three-party and four-party settings are not directly comparable due to the different predicted pie shares. However, the critical point is that in both settings, *Stages1&2* leads to pie shares that are much closer to proportionality than those in *Baseline* or *Stage2*. A random-effects logistic regression (s.e. clustered on matching groups) with dependent variable "distance to proportionality" and the six treatments as independent variables confirms that the differences between *Stage2* and the other two treatments are highly significant for the three-party and the four-party setting (for all four comparisons, $p < .001$).

B.2 Joint test of hypotheses

Results 1 and 2 confirm Hypothesis 1. Result 3 confirms Hypothesis 2. Here, we evaluate the *joint hypothesis* requiring theory to simultaneously explain winning coalitions (Hypothesis 1) and pie shares (Hypothesis 2).

We create the variables Joint Proportional Solution (JPS) and Joint Main Simple Solution (JMSS). JPS is equal to the pie shares predicted by the proportional solution only if an LWC forms. It is equal to 0 otherwise. Likewise, JMSS is equal to the pie shares predicted by the main simple solution only if an MWC forms and is equal to 0 otherwise. These variables can thus explain an empirical outcome only if they are accurate for the winning coalition *and* the pie shares simultaneously.

We run analogous regressions to Table 4, except that we use JPS and JMSS instead of PS and MSS. We present the results in Table 6. The random-effects regressions examine how proposers' (model 1) and acceptors' (model 2) pie shares depend on JPS and JMSS for the different negotiation environments. Consistent with our main findings, JPS performs best at explaining outcomes in *Stages1&2* while JMSS performs best in *Baseline* and *Stage2*. The coefficients in Table 6 are smaller than the ones reported in Table 4 because, by definition, fewer negotiation outcomes are consistent with JPS/JMSS than PS/MSS.

Table 6: Proportional and Main Simple Solution—Joint Test

	(1) Proposer pie share		(2) Acceptor pie share	
Stage2	-4.326*	(2.553)	3.897*	(2.346)
Stages1&2	5.000*	(2.625)	2.089	(2.845)
Baseline × JPS	0.0579	(0.0452)	0.137**	(0.0618)
Stage2 × JPS	0.0743**	(0.0300)	0.0633*	(0.0385)
Stages1&2 × JPS	0.193***	(0.0327)	0.202***	(0.0187)
Baseline × JMSS	0.117**	(0.0551)	0.283***	(0.0829)
Stage2 × JMSS	0.204***	(0.0625)	0.241***	(0.0389)
Stages1&2 × JMSS	-0.0685	(0.0608)	0.148**	(0.0607)
Constant	25.45***	(1.750)	14.04***	(1.753)
Wald tests comparing effect of JPS/JMSS across commitment setting				
Stages1&2 × JPS = Baseline × JPS	$p = 0.016$		$p = 0.303$	
Stages1&2 × JPS = Stage2 × JPS	$p = 0.008$		$p = 0.002$	
Baseline × JPS = Stage2 × JPS	$p = 0.768$		$p = 0.283$	
Stages1&2 × JMSS = Baseline × JMSS	$p = 0.020$		$p = 0.174$	
Stages1&2 × JMSS = Stage2 × JMSS	$p = 0.001$		$p = 0.125$	
Baseline × JMSS = Stage2 × JMSS	$p = 0.294$		$p = 0.630$	
Wald tests comparing effect of JPS/JMSS within commitment setting				
Baseline × JPS = Baseline × JMSS	$p = 0.537$		$p = 0.306$	
Stage2 × JPS = Stage2 × JMSS	$p = 0.120$		$p = 0.004$	
Stages1&2 × JPS = Stages1&2 × JMSS	$p = 0.001$		$p = 0.392$	
Period dummies	✓		✓	
Party size dummies	✓		✓	
Three-party/four-party dummies	✓		✓	
Negotiations (N)	710		783	
Unique representatives (subjects)	327		334	
Matching groups (clusters)	24		24	

Notes: * significant at 10%, ** significant at 5%, *** significant at 1%. Random effects regressions (individual and matching group random effects) with standard errors in parentheses are clustered on matching groups. All regressions include dummies for the size of a party and whether the observation stems from a three-party or four-party treatment to improve the fit for non-MWC winning coalitions for which JPS and JMSS are equal to 0. Reference group: proposers (model 1) or acceptors (model 2) in *Baseline*.