Power illusion in coalitional bargaining: An experimental analysis

Nicola Maaser
Dept. of Economics and Business Economics, Aarhus University
[*Corresponding Author*]

Fabian Paetzel
Dept. of Economics & FOR 2104, Helmut-Schmidt University Hamburg

Stefan Traub
Dept. of Economics & FOR 2104, Helmut-Schmidt University Hamburg

July 29, 2019

Abstract

One feature of legislative bargaining in naturally occurring settings is that the distribution of seats or voting weights often does not accurately reflect bargaining power. Game-theoretic predictions about payoffs and coalition formation are insensitive to nominal differences in vote distributions and instead only depend on pivotality. We conduct an experimental test of the classical Baron-Ferejohn model with five-player groups. Holding real power constant, we compare treatments with differences in nominal power. We find that initial effects of nominal differences become small or disappear with experience. Our results also point to the complexity of the environment as having a negative impact on the speed at which this transition takes place. Finally, and of particular importance as a methodological observation, giving subjects a pause accelerates learning.

Keywords: legislative bargaining, alternating offers, experiments, weighted voting, coalition formation

JEL codes: C78, C92, D71, D72, C7, C52

We thank the associate editor and two referees for their constructive comments. We are grateful to Guillaume Fréchette for making the original data from Fréchette et al. (2005a) available. We also thank André Bächtiger and Stefan Napel for helpful comments on earlier drafts. We benefited from feedback on seminar presentations in Aarhus, Bayreuth, Siegen, at the 2017 meeting of the Verein für Socialpolitik, the ESA European Meeting in Vienna 2017, and the GfEW meeting 2017 in Kassel. Financial support from the University of Bremen Central Research Development Fund and the German Research Foundation (Stefan Traub, reference TR 458/6-1) is gratefully acknowledged.

1 Introduction

Collective decision-making frequently involves situations in which actors have different numbers of votes. Various political institutions assign heterogeneous voting weights explicitly. Examples include the Council of the European Union, the Board of Governors of the International Monetary Fund, and the U.S. Electoral College. Choice of weights sparks recurrent controversies in such bodies.[1] More generally, weighted voting arises when votes are combined and cast together. Important examples are the formation of coalition governments, voting in legislatures with cohesive factions, and shareholder voting in corporations.

The combinatorial nature of weighted voting implies that weights need to be distinguished from power. Seemingly different games can be strategically equivalent. For illustration, suppose that elections to a 100-seat simple-majority legislature result in five parties winning seats, and the seat distribution is $(42, 33, 9, 9, 7)$. On closer examination, this seat distribution is isomorphic in terms of parties' possibilities to form winning coalitions to $(2, 1, 1, 1, 0)$, with a simple majority threshold of 3. These weights now readily expose that the largest party can form a winning coalition with any of the three middle parties, that all three of the middle parties must combine to exclude the largest, and that adding the seats of the smallest party never turns a losing coalition into a winning one. A remarkable real-world case where the distinction between weights and power was apparently overlooked is presented by the early European Economic Community: Under the voting rules in use between 1958 and 1973, Luxembourg's vote could not sway the outcome of any division.[2]

The present paper experimentally investigates how variation of *nominal* power, holding *real* power constant, affects bargaining behavior and outcomes in the influential multilateral bargaining model proposed by Baron and Ferejohn (1989). Two decision-making rules are equivalent (imply the same real power) if both generate identical sets of winning coalitions. Any differences between them that do not alter the set of winning coalitions are nominal. In the game under consideration, players hold equal real bargaining power while nominal power may differ. Nominal power differentials are, from the perspective of standard non-cooperative game theory, extraneous to the distribution of equilibrium payoffs and to coalition formation. We refer to the converse idea that nominal differences might matter as *power illusion*.

The effect of nominal power differentials is worth studying for at least two reasons: First, numerous empirical analyses of coalition governments find that payoffs to coalition members like, e.g., ministerial posts, bear an almost proportional relationship to the nominal votes each coalition partner contributes to the coalition (see Warwick and Druckman 2006). These studies seem to support "Gamson's Law", a conjecture named after William Gamson, who asserted that "Any participant will expect others to demand from a coalition a share of the payoff proportional to the amount of resources which they contribute to a

[1] The EU's negotiations of the Treaty of Lisbon saw a particularly heated debate on voting rules. See, e.g., *The Economist* (2007, June 14th).

[2] In the parlance of cooperative game theory, Luxembourg in the Council of Ministers and the smallest party in the above example were *null players*. However, Mayer (2018) gives a more nuanced assessment of Luxembourg's voting power in the early European Economic Community.

coalition" (Gamson 1961, p. 361). By contrast, non-cooperative theory predicts that only *real* differences as captured by the set of winning coalitions will be relevant (see Snyder et al. 2005 and Montero 2006).[3] This contradiction is discussed from an empirical perspective by Cutler et al. (2016).[4] Second, most bargaining or voting situations in naturally occuring settings are highly complex, and those involved might hence rely on nominal weights as cognitive short-cuts. Therefore, we should study whether this is already the case in environments that are relatively easy to understand.

We conduct an experiment where we compare three nominally different representations of simple majority rule in a Baron-Ferejohn game with five players. Our results show that, while unexperienced players significantly respond to the framing of the voting rule, effects are mostly weak or disappear for experienced players. We study the effects of experience in more detail by holding (i) experiments with 20 consecutive repetitions of the bargaining situation and (ii) experiments that involved a short break between two sets of ten repetitions. A second main finding is that, surprisingly, the influence of nominal weighting persisted to some extent in uninterrupted treatments. For example, payoffs exhibited, as a consequence of behavioral changes, a considerable degree of heterogeneity between players that are identical from a theoretical perspective. By contrast, illusionary behavior virtually disappeared when players had a break.

To the best of our knowledge, the only other study of the question whether nominal power will matter is Fréchette et al. (2005a). It includes two treatments, which are nominally different variants of three-player simple majority rule in a Baron-Ferejohn game. They find only very minor and transitory differences between treatments and conclude that nominal voting power has no significant impact. Our design allows us to trace this assessment largely back to the 'break effect'.

The remainder of the paper is organized as follows: Section 2 describes the Baron-Ferejohn model and relates our study to previous experimental literature. Section 4 reports the experimental results. We discuss the effect of the break and of the complexity of the bargaining situation in Section 5. Section 6 concludes. Additional materials are presented in three appendices.

2 The bargaining game

2.1 Overview

Consider a committee comprising n members who decide on how to split a fixed budget normalized to 1. The committee uses a weighted voting rule $[q; w_1, \ldots, w_n]$ where w_i

[3]Proposition 5 in Montero (2006) applies to all constant-sum homogeneous games with any number of players, while results in Snyder et al. (2005) and Ansolabehere et al. (2005) are restricted to large replicated games. The latter authors offer an application to portfolio distribution.

[4]Specifically, Cutler et al. (2016) study government coalitions in European parliamentary democracies and find that, controlling for raw seat shares, parties with higher real (minimum integer) weights are more likely to be members of the government coalition. Yet, they find nominal weights ('raw' seats) to be better at predicting the portfolio allocation conditional on coalition membership.

is legislator i's voting weight, and q is the quota. A coalition $S \subseteq N$ is *winning* iff $\sum_{i \in S} w_i \geq q$; it is *minimal winning* if it is winning and no $T \subsetneq S$ is winning.

The Baron-Ferejohn model portrays this situation as a sequential non-cooperative bargaining game over multiple periods. It has been used in a wide range of applications, and it has been extended in several directions, e.g., special interest politics, the formation of coalition governments, and the geographic distribution of public expenditures.[5] In all variants of the model, a proposer is selected randomly according to a known recognition *protocol*. The two most common assumptions are that recognition probabilities are proportional to players' voting weights, or that recognition probabilities are all equal.[6] The proposer then puts forward an allocation of resources $(x_1, \ldots, x_n) \in \mathbb{R}^n_+$ to the other players, subject to not exceeding the total budget constraint, i.e., $\sum_i x_i \leq 1$. Next, players simultaneously vote yea or nay, and the proposal is then either accepted or rejected according to the weighted voting rule. In the closed rule version of the game, no member of the committee can offer an amendment. If the proposal fails, then a new proposer is selected at random with the process being repeated until an allocation is determined.

We focus on the closed rule Baron-Ferejohn game with five players and simple majority rule. Thus, any coalition which includes at least three players can decide how to divide the pie. The canonical *representation* of this voting rule is $R = [3; 1, 1, 1, 1, 1]$, i.e., each agent has one vote, and three votes are needed to pass a proposal.[7] R reflects the real bargaining power of the players in a particularly transparent way. We will therefore use R as a baseline in our experiment. Each weighted voting rule, however, has an infinite number of other representations, some of which can be obtained, for example, by multiplying weights and quota by the same positive constant. But in addition, other representations exist that lead to the same possibilities for coalition formation and hence leave the theoretical analysis unaffected. Our experimental design involves two other representations of the canonical voting game. We chose, first, the representation $R' = [7; 2, 2, 3, 3, 3]$ which might be seen as more complex than R as it involves two 'types' of players (weight-2 and weight-3) rather than one. Second, we consider $R'' = [18; 5, 6, 7, 8, 9]$ because it has the smallest integer weights such that the weights of all five players differ.

In order to allow for a clean assessment of the potential effect of such a non-transparent representation, our experimental implementation focuses, first, on the equal recognition protocol, i.e., in each round each player has probability $p_i = 1/5$ to be the proposer. Unlike other protocols, equal recognition avoids any theoretical differences between the three representations of the voting rule. Second, we assume that the budget does not shrink if the proposal does not pass. This ensures that the potential exclusion of some players from the winning coalition – and not discounting of future payoffs – is the key

[5] Eraslan and McLennan (2013) provide a review of the literature.

[6] The latter is also consistent with an institution such as the Council of the European Union, in which countries' votes count differently, but opportunities for making proposals are, at least formally, equal.

[7] In cooperative game theory, a weighted voting rule is called *homogenous* if it admits a representation $[q^h; w_1^h, \ldots, w_n^h]$ such that all minimal winning coalitions have exactly the same total voting weight. Here, R is a homogeneous minimum integer representation (see Freixas and Kurz 2014). – Note that not all weighted voting rules allow homogeneous representations, nor are homogeneous representations necessarily unique.

determinant of the equilibrium.

As is standard in the literature, we restrict attention to subgame perfect equilibria in stationary strategies (SSPE) as a benchmark prediction.[8] In such an equilibrium, the proposer offers amounts to her proposed coalition partners that make them exactly indifferent between voting for and against (their continuation values), and she keeps the residual for herself. Voters' fear of being excluded in future proposals affords a great advantage to the proposer, even in the absence of time discounting. In our specific five-player game the outcome prediction of the unique SSPE is that the proposer allocates 3/5 of the pie to herself, and the continuation value of 1/5 to each of two other agents. Two agents are left out of the minimal winning coalition and receive nothing. The proposal is approved without delay. Ex ante, i.e., before a proposer has been selected, expected payoffs equal, of course, 1/5 for each player.

2.2 Related literature

Dating back to McKelvey (1991), there have been a number of experiments that have aimed at testing the Baron-Ferejohn model. Generally, several features of the game's stationary equilibrium are qualitatively supported by the laboratory results, namely, infrequent delay and a high frequency of minimal winning coalitions. Moreover, Fréchette and Vespa (2017) find that the continuation value is indeed a main determinant of voting decisions.

Observed proposer power is not nearly as strong as theoretically predicted; proposers typically enjoy some advantage over other members of the coalition, but their claims mostly fall short of the SSPE.[9] Generally, play in experimental bargaining games consistently deviates from the SSPE predictions in favor of a more equal distribution of benefits between coalition members (see, e.g., Fréchette et al. 2003).[10] A recent study by Nunnari and Zapal (2016) reexamines data from several experiments on Baron-Ferejohn bargaining. It explains the observed deviations from the theoretical predictions in terms of imperfect best responses and players' incorrect beliefs about their chance to become proposer in the future.

As already mentioned above, only the article by Fréchette et al. (2005a) has introduced nominal weights into the Baron-Ferejohn model to test real versus nominal voting power. The element of their study that is of particular interest in our context is the comparison between representations $[50; 33, 33, 33]$ and $[50; 45, 45, 9]$, referred to as "equal-weight-equal-selection" (EWES) and "unequal-weight-equal-selection" (UWES), respectively. The authors find that, in line with game theoretic bargaining models, the nominal change in

[8] Stationarity rules out any dependence of the agents' strategies on the history of play. Baron and Ferejohn (1989) argue that, while any outcome can be supported as a subgame perfect equilibrium if players are sufficiently patient, the stationary equilibrium is a focal point.

[9] However, Agranov and Tergiman (2014) and Baranski and Kagel (2015) show that allowing subjects to engage in (cheap-talk) communication before a proposal is submitted brings experimental results much closer to the theoretical prediction.

[10] Fréchette (2009) provides a re-analysis of the data from Fréchette et al. (2003) in the light of a belief-based learning model.

the number of votes between EWES and UWES had no effect on key outcomes such as proposer's self-offer or voting behavior. Moreover, the composition of winning coalitions was not found to be biased in favor of players with weak nominal bargaining power. We discuss this result further in Section 5.2.

Fréchette et al. (2005b) compare the predictions of the Baron-Ferejohn model and a demand bargaining model (see Morelli 1999) with Gamson's Law in a three-player majority setting. The study draws on the same data as Fréchette et al. (2005a) with regard to the Baron-Ferejohn game; thus, the conclusions about nominal power are identical. Fréchette et al. (2005c) report on Baron-Ferejohn and demand bargaining with five players in both an equal weights treatment, i.e., $[3; 1, 1, 1, 1, 1]$, and in the apex treatment $[4; 3, 1, 1, 1, 1]$. Here, one player has disproportionate real voting power so that this study is not a test of nominal versus real voting power. Diermeier and Morton (2005) investigated a finitely repeated variant with three-member groups and varied, one, the share of votes that each subject controlled, and, two, subjects' recognition probabilities. In all treatments there was a total of 99 votes, and a threshold of 50 votes for approval. Vote assignments in their three treatments were (34, 33, 32), (49, 33, 17) and (46, 44, 9), respectively. All three are equivalent to the situation where each player has one vote, and a coalition of at least two players can pass a proposal. Still, in contrast to our paper, Diermeier and Morton (2005) do not study purely nominal power differentials. The reason is that they base the probability of being selected as a proposer on the subject's percentage of votes, so that theoretical predictions, especially regarding choice of coalition partners, differ across their treatments. The authors find little support for the predictions of the model as proposers allocate money to all players rather frequently and do not exploit their proposal power. But neither did a proportionality norm based on the vote share explain the experimental data well. Diermeier and Morton conclude that subjects' behavior is best accounted for by an equal sharing rule, where the proposer chooses any winning coalition and then distributes payoffs equally among the coalition members. Drouvelis et al. (2010) conduct an experiment where the primary motivation is to test the hypothesis that adding a new player to a weighted voting game could increase the voting power of an original player. Yet, one of their treatments uses the three-person simple majority game $[4; 3, 2, 2]$, where one player is nominally stronger than the other two. Results from this treatment provide no evidence for the existence of power illusion. For example, differences in earnings or in the acceptance behavior between the 'strong' and the 'weak' players are insignificant.

Furthermore, this work is related to other studies that have introduced asymmetries into the Baron-Ferejohn framework. In theoretical work, Snyder et al. (2005) extend the Baron-Ferejohn framework to accommodate general weighted voting rules. Excluding situations where real and nominal power may diverge, they use large replicated games to show that voters' expected payoffs are – in the limit – proportional to voting weights, at least under the assumption of proportional recognition probabilities. However, this result does not necessarily carry over to the small games usually found in applications (see Montero 2017). Vespa (2016) analyzes in a Baron-Ferejohn experiment how the apportionment of proposal power and voting power affects the distribution of resources under weighted voting or bicameralism. Papers studying other aspects of committee bargaining include: Miller and

Vanberg (2015) (effect of group size), Kagel et al. (2010) (add veto rights) and Miller et al. (2015) (add heterogeneity in disagreement values).

3 The experiment

3.1 Experimental procedures

The experiment comprised 12 sessions conducted at the University of Hamburg.[11] Participants were undergraduate and graduate students from various disciplines, recruited from an extensive subject pool via the administration software h-root (Bock et al. 2014). Each session included 20 bargaining periods. We implemented a 3×2 treatment design, where voting weights nominally differ in three ways (referred to as BASE, PIT1, PIT2), and where the 20 periods were either played consecutively (NO BREAK) or with a short break of ten minutes after the first 10 periods (BREAK). Experimental treatments are summarized in Table 1.

In each bargaining period of an experimental session groups of five subjects had to split 150 tokens among themselves by simple majority rule. In the baseline treatments (BASE) subjects were each given a voting weight of 1. Subjects were assigned voting weights $\{2, 2, 3, 3, 3\}$ in PIT1 and $\{5, 6, 7, 8, 9\}$ in PIT2. Thus, nominal weights created one, two or five 'types' of players in the three representations. We conducted two sessions for each of the resulting six treatments, involving 25 subjects per session. Thus, 300 subjects participated in total; Table A1 in the Appendix summarizes subjects' socio-demographic characteristics.

To minimize repeated game effects, subjects were randomly rematched into groups between periods, but not between the rounds within a given bargaining game, without having the possibility to learn others' identity. In PIT1 matching was subject to the restrictions that each five member group contained two players of weight 2 and three players with weight 3, and (ii) that each subject assumed weight-2 type (weight-3 type) four (six) times over the course of the first ten and the last ten periods, respectively. Similarly, matching in PIT2 required (i) that each five member group contained one player of each weight type, and (ii) that each subject assumed each of the five weight types twice over the course of the first ten periods and twice over the course of the last ten periods.

Within a given period, subjects were identified by a label (a symbol) that remained fixed for the duration of that period. The sequence of events was as follows: First, each subject was randomly assigned her voting weight. Each subjects got to know her own voting weight as well as the voting weights of the other four group members and the quota. Each player was prompted to enter her proposal on how to allocate the 150 Tokens, i.e., to

[11] The experiment was programmed using z-tree (Fischbacher 2007) and took place between February and March 2018. Subjects were given copies of the instructions (see Online Appendix); all questions were answered in private. All sessions lasted about 120 minutes including instruction time. We earlier ran a similar experiment that included only one treatment (now PIT2) besides the control and had a time horizon of only ten periods. Data are available from the authors upon request.

Table 1. Treatment Conditions, Sessions and Subjects

		NO Break (20 periods)	Break (10 + 10 periods)
BASEline treatments	$[3; 1, 1, 1, 1, 1]$	BASE-NOB 2 sessions 25 + 25 subjects	BASE-B 2 sessions 25 + 25 subjects
Power Illusion Treatments 1	$[7; 3, 3, 3, 2, 2]$	PIT1-NOB 2 sessions 25 + 25 subjects	PIT1-B 2 sessions 25 + 25 subjects
Power Illusion Treatments 2	$[18; 9, 8, 7, 6, 5]$	PIT2-NOB 2 sessions 25 + 25 subjects	PIT2-B 2 sessions 25 + 25 subjects

specify a vector $(x_1, \ldots, x_5)$ of integer token amounts, with no waste allowed. Others' labels and voting weights were displayed in randomized order to avoid anchoring effects. One of the five proposals was randomly chosen (with probability 1/5 for each) and displayed to all group members. The subject whose proposal was selected was not given the possibility to vote upon it; her approval was taken as granted. The remaining four players simultaneously voted in favor or against the standing proposal. If the proposal obtained a simple majority, the proposed distribution became binding, and that period was concluded for the five group members. If the proposal failed, a new round of the same game was initiated, where one player – possibly the same as before – was randomly selected (with probability 1/5) to make a proposal. This was repeated until an allocation was achieved, with no shrinkage in the amount of tokens to be allocated.

Following each vote, detailed feedback was displayed to the subjects within each group. This consisted of the proposed distribution of benefits, the proposer's weight type, the individual votes cast by each player together with their voting weight, and whether the distribution passed or failed. If the proposal passed, subjects were informed about their individual payoff in that game. At the end of a session, one period was randomly chosen, respectively, from the first ten periods and from the last ten periods, to be paid off in private. The exchange rate was 1 token = 0.20 € and subjects also received a 10 € show-up fee. Earnings averaged 22 € ranging from 10 € to 34 €.

3.2 Conjectures

It is well-known that experimental outcomes in alternating offers bargaining typically differ from the theoretical point predictions outlined above. Here we will be interested instead in how the introduction of nominal voting weights affects players' behavior as proposers and responders, and expected payoffs. We use the BASE treatment as a benchmark and consider any deviations that occur in the treatments with nominal asymmetries as 'power illusion'. Since theory asserts that nominal differences should have no effect whatsoever, we refer to our following expectations as conjectures rather than hypotheses.

We suppose that having greater or smaller weight than others might affect proposers' allocations.

Conjecture 1 (Proposers).

(a) *The share that a player allocates to herself increases in her voting weight.*

(b) *A proposer allocates more to a member of the coalition the greater the latter's voting weight.*

The multilateral bargaining model suggests that players will base their decision on whether to accept or reject a proposal solely on the amount offered to them. In particular, they should accept any offer that at least corresponds to their continuation value. Alternatively, individuals' minimal acceptable offer might be influenced by nominal weights.

Conjecture 2 (Responders).

A players' acceptance threshold is higher the larger her weight.

Proposers who maximize their own payoff will choose coalition members based on their expectations about the latters' acceptance thresholds.

Conjecture 3 (Coalitions).

(a) *Players are less often included in others' coalitions the larger their weight.*

(b) *Relative coalition weight is smaller in* Pit1 *and* Pit2 *than in* Base.

Conjecture 3(a) follows as a corollary from Conjecture 2 and Conjecture 3(b) is a direct consequence of 3(a). As compared to Pit1, Pit2 increases the degree of complexity by assigning five different weights to the players.

Conjecture 4 (Variation).

We expect power illusion effects to be stronger in Pit2 *compared to* Pit1.

A common assumption in experimental economics is that subjects learn through repetitions (see, e.g., Loewenstein 1999). Moreover, it has been suggested that players "cool down", that is, make more reflected decisions, if bargaining is interrupted (see, e.g., Grimm and Mengel 2011 and Oechssler et al. 2015 in the context of ultimatum bargaining). This leads us to the following conjecture:

Conjecture 5 (Time).

(a) *Subjects utilize experiences of previous rounds and therefore exhibit less or no power illusion in later periods.*

(b) *We expect power illusion to diminish after subjects are given a short break.*

4 Results

We first present our experimental results regarding subjects' behavior along the structure of Conjectures 1–3 above. Subsection 4.4 looks into players' payoffs. We address the impact of experience (Conjecture 5(a)) in all subsections. Throughout, we focus on results from the first five and the last five periods, in order to distinguish clearly between inexperienced and experienced behavior. Results for Periods 6-10 vs. 11-16 as well as period-by-period results are included in the Online Appendix. The overarching Conjectures 4 and 5(b) are studied in detail in Section 5.

4.1 Proposers' behavior

Figure 1 shows the amount of tokens that proposers allocate to themselves by treatment, own weight, experience and whether or not subjects had a break after the first ten periods. The analysis is based on first round proposals only. All players had to submit a first round proposal before learning their roles either as proposers or responders.

The two left panels of Figure 1 suggest, in line with Conjecture 1(a), that unexperienced proposers in Pit1 and Pit2 tend to offer more to themselves when they were assigned larger weights. The effect is marked for weight-3 players in the Pit1 treatments who claim significantly more tokens than weight-2 proposers (7.51 in Pit1-nob and 5.16 in Pit1-b).[12] Similarly, we find a positive and almost monotonic relationship between an unexperienced proposer's weight and her self-offer in Pit2 treatments. Here, claims by weight-9 proposers are between 7.40 and 8.54 tokens larger than those made by weight-5 proposers. A test on the equality of several medians rejects the null hypothesis that self-offers of *un*experienced proposers are uniform across nominal weights in Pit2-nob and Pit2-b ($\chi^2 = 18.50$, $p \leq 0.01$; and $\chi^2 = 19.84$, $p \leq 0.01$, respectively).

We also note that unexperienced weight-2 and weight-5 proposers claim less for themselves than unexperienced weight-1 proposers in the baseline treatments. For all treatments and weights, unexperienced proposers' claims are significantly lower than predicted by the SSPE. Of course, self-offers falling short of the theoretical benchmark could still reflect expected utility maximization if responders reject proposals close to the SSPE at a high rate. It is a well-established finding in experimental alternating offers games that proposers learn to keep somewhat more for themselves over time, yet nowhere near the very large share predicted (see, e.g., Fréchette et al. 2003).

A glance at the right-hand panels of Figure 1 immediately shows that experience exerted a strong impact on proposers' self-offers. In Pit1-nob (upper right panel), experienced weight-3 players on average kept only 4.76 tokens more for themselves than weight-2 players; if experience was combined with a break (lower right panel), the mean difference was only 1.29 tokens and became insignificant. Likewise, in Pit2 the mean difference between

[12] For numbers of observations, means, standard errors and t-tests see the left panel of Table A2 in the Appendix. Note that we had to delete a total of 40 observations in Periods 16-20 of Pit1-nb and Pit1-b due to a software error. We provide results for the intermediate Periods 6-10 and 11-15 in Figure B.1 and period-by-period results in Figure B.3 in the Online Appendix.

weight-9 and weight-5 proposers' demands shrunk to 2.00 without a break and to 1.54 tokens with a break and became insignificant. The null hypothesis that proposers' demands are uniform is neither rejected in Pit2-nob ($\chi^2 = 6.52$, $p = 0.163$) nor in Pit2-b ($\chi^2 = 2.88$, $p = 0.578$). The distinctly lower χ^2-value in Pit2-b still points to the importance of the break (see Section 5).

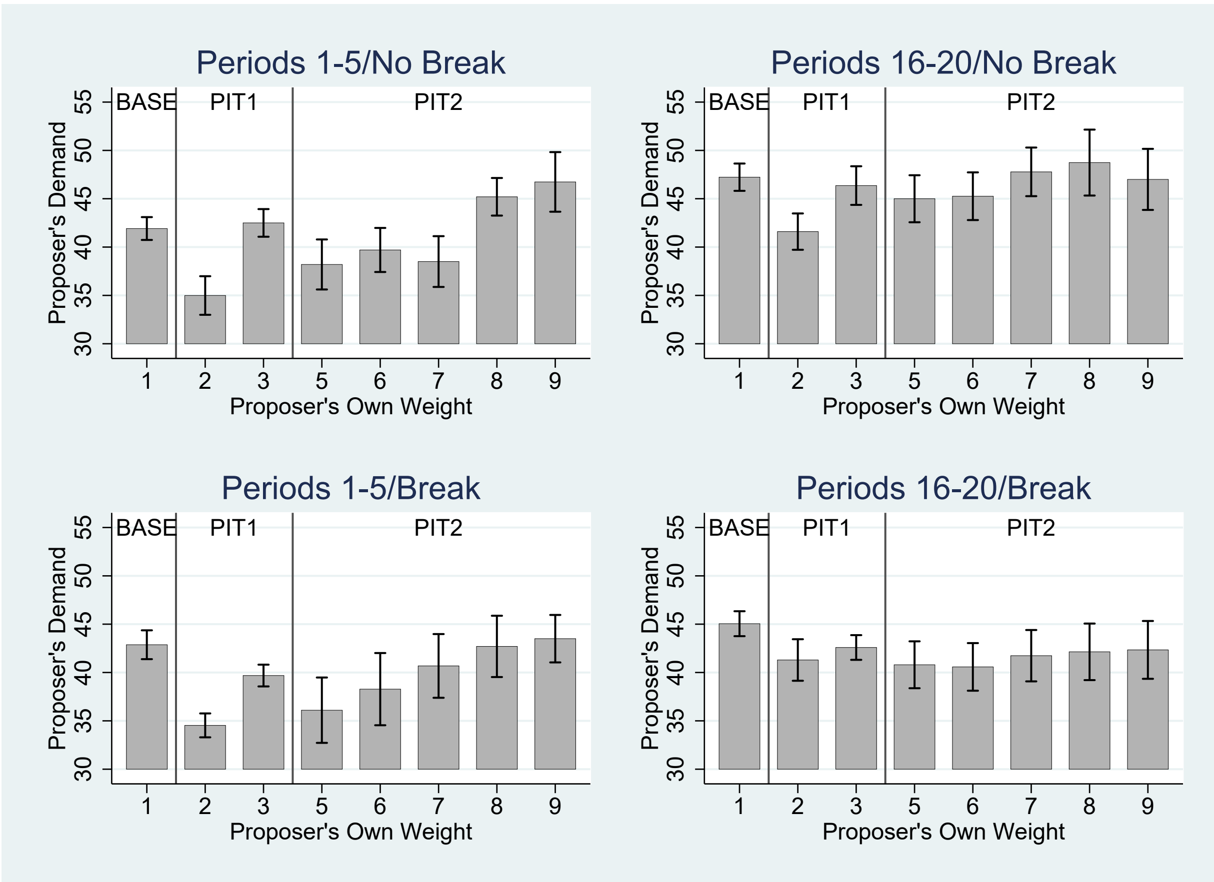

Fig. 1. Proposer's Demand by Proposer's Own Weight, Break and Experience (Periods 1-5 vs. 16-20). The figure uses *all* proposals, selected or not, in the first round of each bargaining period. Error bars represent 90% confidence intervals. For details see the left panel of Table A2 in the Appendix.

The right panel of Table A2 in the Appendix reports predicted values from a GLS random effects model that includes sociodemographic covariates. As should be the case with random assignment of subjects to the different treatments, the predicted values are close to raw data and corroborate the graphical impressions. Note, however, that due to distinctly smaller standard errors the mean differences between experienced weight-9 and weight-5 players remain significant both in Pit2-nob and Pit2-b.

Result 1: *The data support Conjecture 1(a) that the share that a player allocates to herself increases in her nominal voting weight for unexperienced subjects only. Experience* *eliminated this type of power illusion in proposers.*

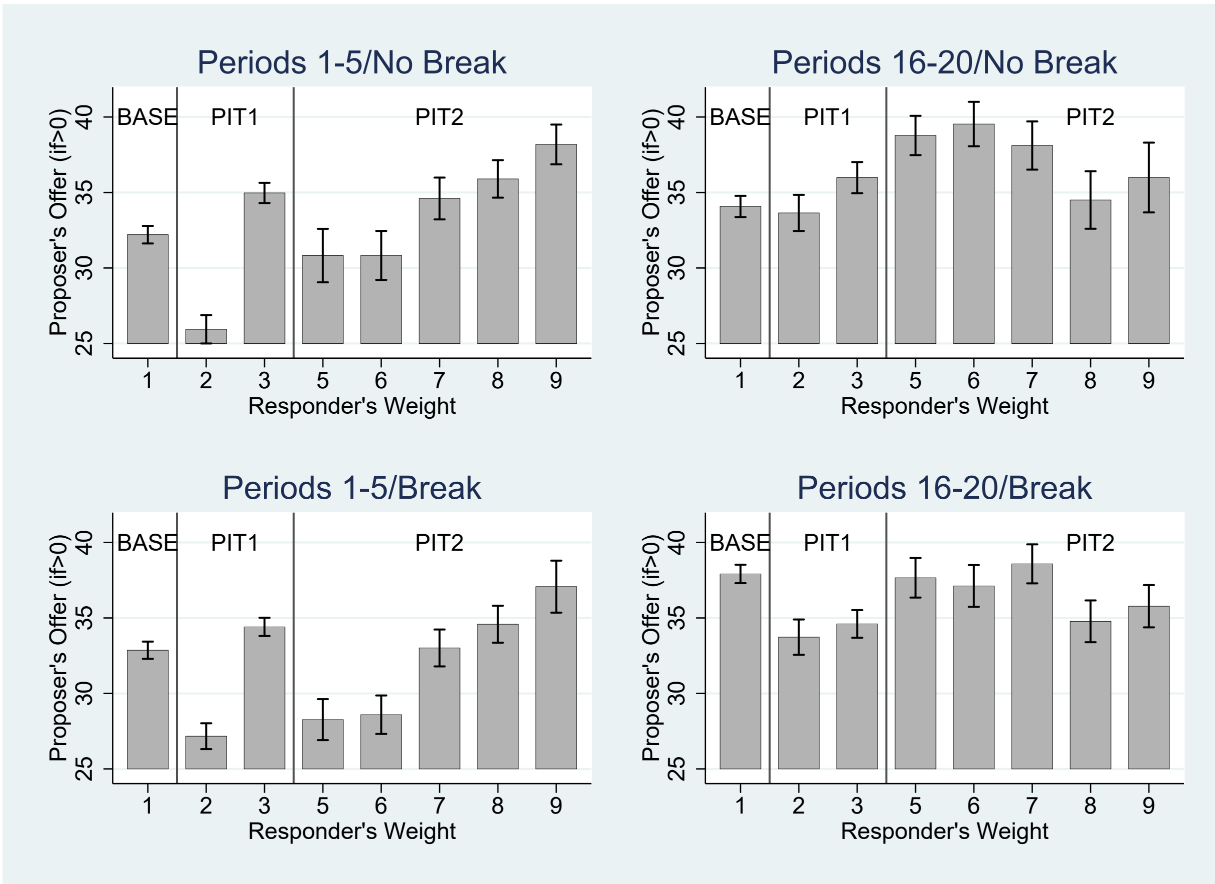

Fig. 2. Proposer's Offer (if > 0) by Responder's Weight, Break and Experience (Periods 1-5 vs. 16-20). The figure uses *all* proposals, selected for voting or not, in the first round of each bargaining period. Error bars represent 90% confidence intervals. For details see the left panel of Table A3 in the Appendix.

We next analyze the non-zero offers that proposers make to others. As with self-offers, we focus on proposals in the first round of each bargaining period. The two left-hand panels of Figure 2 show that unexperienced players (periods 1-5) offer on average more, the more weight a responder wields, supporting our Conjecture 1(b).[13] Offers to the nominally largest and smallest responders diverge greatly, with gaps of 9.03 tokens and 7.24 tokens in PIT1-NOB and in PIT1-B, respectively, and 7.36 tokens and 8.81 tokens in PIT2-NOB and PIT2-B. χ^2-tests reject the null hypothesis that proposers' offers are uniform for *un*experienced players both in PIT2-NOB ($\chi^2 = 43.93$, $p \leq 0.01$) and in PIT2-B ($\chi^2 = 54.30$, $p \leq 0.01$). Unexperienced proposers offer weight-2 and weight-5 players

[13]For numbers of observations, means, standard errors and t-tests see the left panel of Table A3 in the Appendix. We provide results for the intermediate Periods 6-10 and 11-15 as well as period-by-period results in Figures B.6 and B.7 in the Online Appendix.

less, and weight-3 and weight-9 players more, than the average offer to weight-1 players in the Base treatments.

The upper right and lower right panels of Figure 2 indicate that, again, experience evens out differences across weight-types. These amount to only 2.34 tokens and 0.88 tokens in Pit1-nob and Pit1-b, respectively, with the latter mean difference being insignificant. Interestingly, in Pit2 the mean difference between weight-9 and weight-5 responders' offers was negative (insignificant) at -2.79 and -1.88 tokens. The null hypothesis of uniform offers is rejected for experienced players; the respective χ^2-statistics are distinctly lower though ($\chi^2 = 15.19$, $p \leq 0.01$ without break and $\chi^2 = 12.81$, $p = 0.012$ with break).[14]

Result 2: *Both in* Pit1 *and* Pit2 *unexperienced proposers condition their offers on responders' weight in line with Conjecture 1(b), i.e., offers increase in responders' voting weight. Experience strongly diminishes or, when combined with a break after the tenth period, eliminates this type of power illusion in* Pit1. *Experienced offers in* Pit2 *run counter to Conjecture 1(b) as players tend to be offered more the smaller their weight.*

4.2 Responders' voting behavior

We expect responders to apply higher thresholds for accepting an offer the greater their weight (Conjecture 2). Yet, players' acceptance thresholds are not directly observed in our experiment.[15] As an alternative approach, we estimate binary regression models in which the dependent variable is the decision to vote 'yes' when offered a certain amount. Figures 3 and 4 depict a responder's probability of voting 'yes' as a function of the share offered to her. The figures contain error bars for the probability of voting 'yes' when a responder was offered $\{0, 10, \cdots, 50\}$ tokens and the respective significance levels of the marginal effect of the responder's weight on acceptance. Additionally, statistical test results (χ^2-value and significance level) of jointly testing all contrasts are stated in the southeast corner of each panel.[16]

We observe that, in Pit1, weight-3 players are significantly less likely to accept an offer than weight-2 players, with the exception of experienced subjects after the break. Table A4 in the Appendix also shows whether 'fair' offers of 30 tokens were accepted. This analysis confirms that unexperienced weight-3 responders were, irrespective of whether a break was announced, about 12 percentage points less likely to accept the offer compared to unexperienced weight-2 responders. This differential almost doubled for experienced players in the treatment without break, but in Pit1-b, experienced weight-2 responders were less willing than weight-3 responders to accept the offer.

[14] The right panel of Table A3 in the Appendix reports predicted values from a GLS random effects model that includes sociodemographic covariates. The predicted values are close to the raw data and corroborate the graphical impressions. The only mentionable difference is that the predicted mean difference between offers to weight-9 and weight-5 players in Pit2-nob is smaller (-0.71) and insignificant.

[15] We could have asked subjects to state an acceptance threshold. However, this approach forces subjects to use a cutoff strategy that depends *only* on their own payoff.

[16] Table A4 in the Appendix provides the numbers (contrasts) and details about the regressions. Figures B.8 and B.9 in the Online Appendix show the results for Periods 6-10 and 11-15.

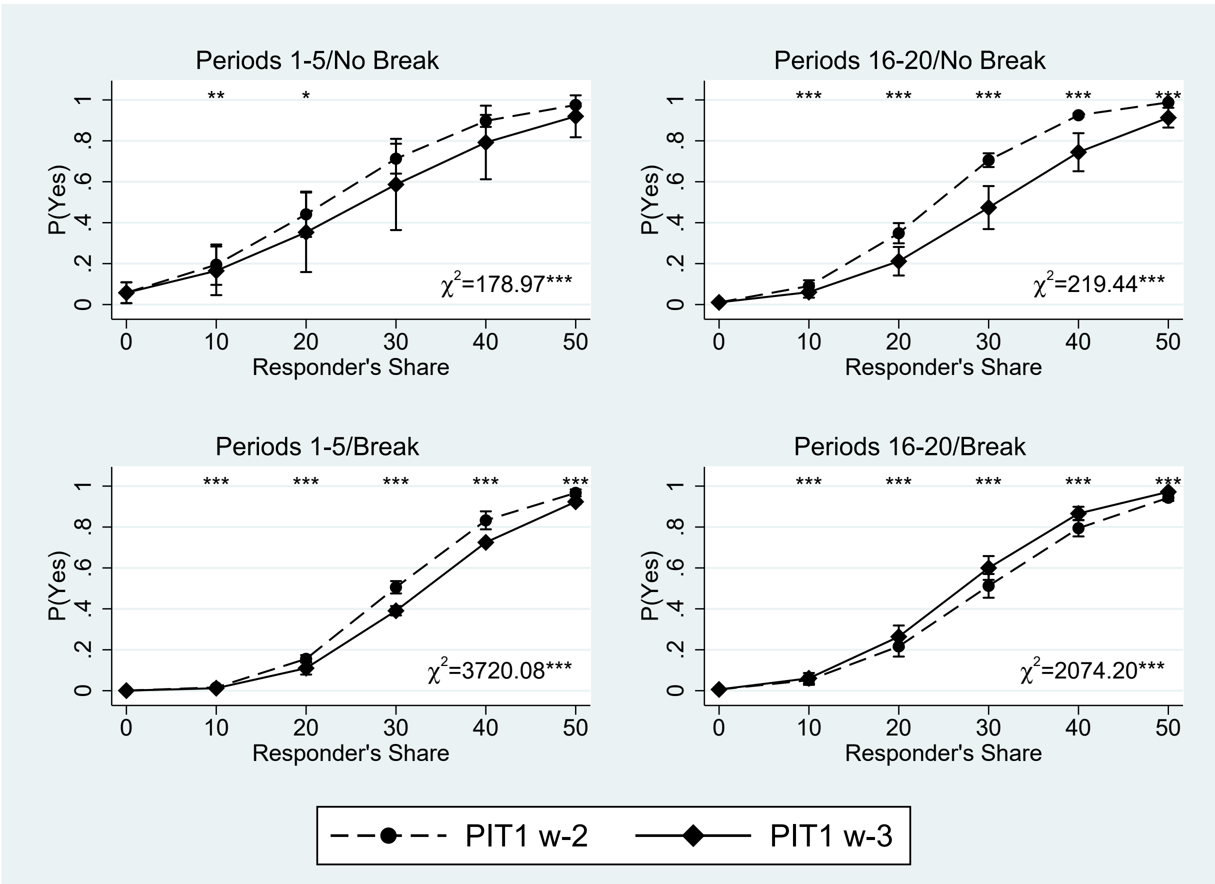

Fig. 3. PIT1: Responder's Probability of Voting 'Yes' as a Function of Responder's Share by Responder's Weight, Break and Experience (Periods 1-5 vs. 16-20). Marginal effects (contrasts) from a random effects Probit panel model with session clustered standard errors. $n = 1,000$ offers. Error bars represent 90% confidence intervals. Asterisks indicate significance levels of contrasts at $\{10, 20, \ldots, 50\}$; χ^2-values for jointly testing all contrasts. Significance levels: $^{***}p \leq 0.01$, $^{**}p \leq 0.05$, $^{*}p \leq 0.1$. For details see Table A4 in the Appendix.

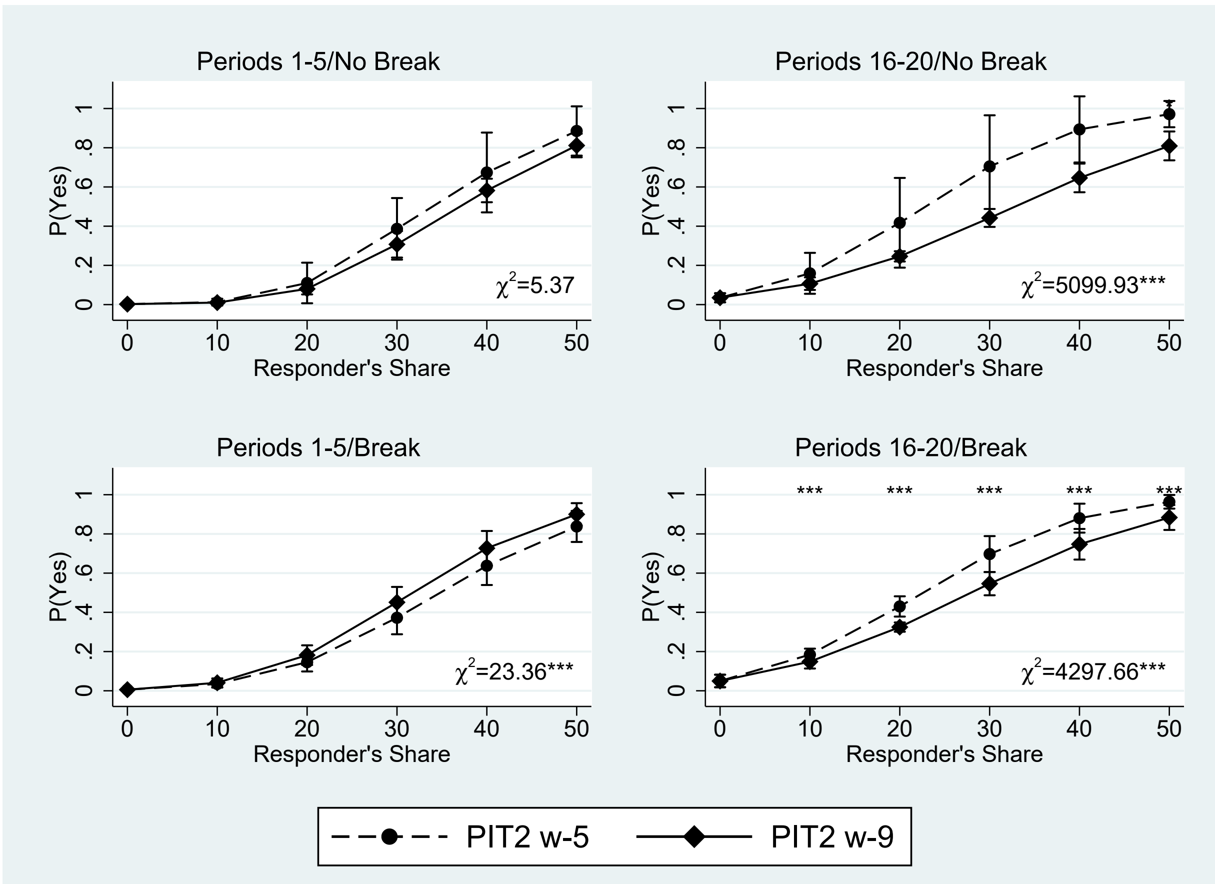

Fig. 4. PIT2: Responder's Probability of Voting 'Yes' as a Function of Responder's Share by Responder's Weight, Break and Experience (Periods 1-5 vs. 16-20). Marginal effects (contrasts) from a random effects Probit panel model with session clustered standard errors. $n = 1,000$ offers. Error bars represent 90% confidence intervals. Asterisks indicate significance levels of contrasts at $\{10, 20, \ldots, 50\}$; χ^2-values for jointly testing all contrasts. Significance levels: $^{***}p \leq 0.01$, $^{**}p \leq 0.05$, $^{*}p \leq 0.1$. For details see Table A4 in the Appendix.

For the sake of clarity, we focus again on the nominally weakest and strongest players in PIT2, rather than including all five weight 'types'. In PIT2-NOB, we do not find any difference in acceptance behavior of unexperienced weight-9 and weight-5 responders. Experienced weight-9 responders, however, are much less likely to accept offers. In the treatment with break, unexperienced weight-9 responders seem slightly more inclined to accept offers than weight-5 responders. After the break, they are significantly less likely to accept offers both in general and for all individual contrasts.

Result 3: *The evidence regarding Conjecture 2 is mixed: A positive relation between a responders' threshold for accepting an offer and their nominal weight is weakly present for unexperienced subjects in both* PIT1-NOB *and* PIT2-NOB*, and gets stronger for experienced subjects. In the break treatments, the effects are less clear as they go in both directions, but are small overall.*

4.3 Coalition formation

In this subsection, we investigate the impact of nominal asymmetries on coalition formation. Conjecture 3(a) suggested that the nominal asymmetries in PIT1 and PIT2 might affect the composition of coalitions by inducing proposers to discriminate against large-weight responders. Figure 5 illustrates the probability of receiving an offer ≥ 30 tokens by treatment and responder's voting weight.[17] The focus here is on offers of 30 or more tokens because SSPE predicts that responders should accept those.

With regard to unexperienced players, Figure 5 illustrates that nominal weights strongly impact on the probability of receiving an offer of at least 30 tokens. In PIT1-NOB and PIT1-B the difference between weight-3 and weight-2 responders is 42 and 36 percentage points, respectively. Compared to weight-1 responders, unexperienced weight-3 responders are included into the coalition more, and unexperienced weight-2 responders less often.

Similarly, the difference between unexperienced weight-9 and unexperienced weight-5 responders is, respectively, 27 and 36 percentage points in PIT2-NOB and PIT2-B. Weight-9 responders receive more, and unexperienced weight-5 responders less, offers ≥ 30 tokens than weight-1 responders do in the baseline treatment.

The two right-hand panels in Figure 5 demonstrate that experience has strong effects.[18] In PIT1, the difference between experienced weight-3 and weight-2 responders drops to 13 percentage points (without break) and 6 percentage points (with break), but remains significant. In PIT2, the difference is turned around: Now, weight-9 responders exhibit a 33 percentage points (without break) and 11 percentage points (with break) lower probability of receiving an acceptable offer compared to weight-5 players. The null hypothesis of uniform probabilities to receive an offer ≥ 30 has to be rejected both for unexperienced players (PIT2-NOB: $\chi^2 = 61.40$, $p \leq 0.01$; PIT2-B: $\chi^2 = 110.40$, $p \leq 0.01$) and for experienced players (PIT2-NOB: $\chi^2 = 79.24$, $p \leq 0.01$; PIT2-B: $\chi^2 = 22.72$, $p \leq 0.01$).

[17] Table A5 in the Appendix contains means, standard errors, and tests based on the raw data as well as the marginal effects of a Probit random effects panel model with session clustered standard errors.

[18] We provide results for the intermediate Periods 6-10 and 11-15 in Figure B.10 in the Online Appendix.

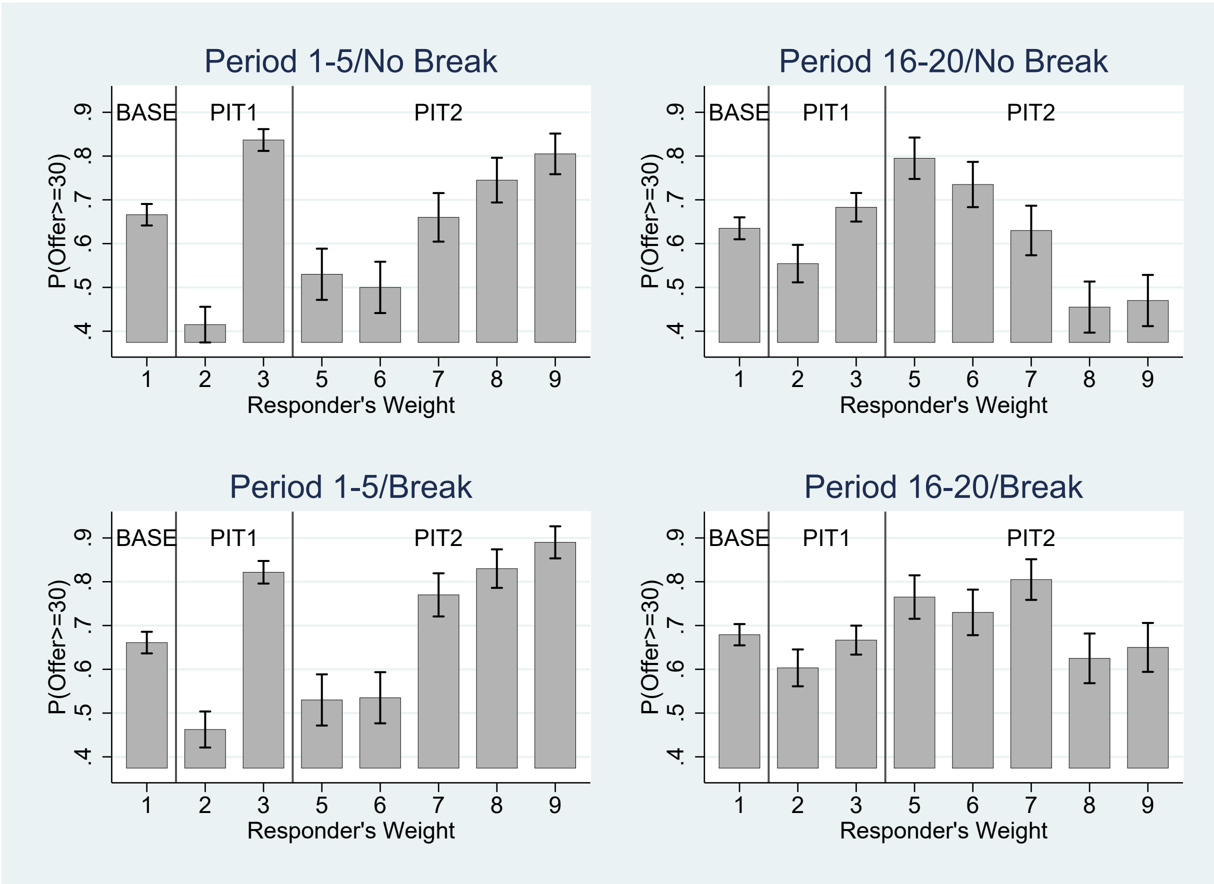

Fig. 5. Responder's Probability of Receiving an Offer$\geq$ 30 by Responder's Weight, Break and Experience (Periods 1-5 vs. 16-20). The figure uses *all* proposals, selected or not, in the first round of each bargaining period. Error bars represent 90% confidence intervals. The number of observations for each bar can be taken from the left panel of Table A5 in the Appendix.

Result 4: *Conjecture 3(a) that players are less often included in others' coalitions the larger their weight is not confirmed by our data. In fact, unexperienced large-weight players are preferred coalition partners. In* Pit1*, experience then greatly reduces the impact of nominal weights on coalition formation. In* Pit2*, by contrast, experienced proposers do discriminate against large weight responders.*

To the extent that small-weight responders are sought after as coalition partners, relative coalition weight, i.e., the combined voting weight of the coalition members as a share of the sum of all weights, will be smaller when nominal asymmetries exist compared to Base treatments (see Conjecture 3(b)). Table 2 reports relative coalition weight in Panel A. Relative coalition weight is significantly smaller only when comparing experienced players in Pit2-nob to corresponding Base outcomes (mean -0.06, standard error 0.03, $p = 0.045$). Other results do not support the conjecture.

Theory predicts that proposers will seek to build minimum winning coalitions by making an offer to any two of the other four players, amounting to 60% in terms of relative coalition weight. Yet, it is not unusual in experimental bargaining that proposers include more players than necessary into their coalition. Panel B of Table 2 shows that absolute coalition size in our experiment also deviates from the SSPE prediction of three players in all cases. Still, on average, 62.8% of all coalitions are minimal, with little variation across different treatments.[19] Theory also predicts that proposals should get accepted without delay. In line with other experimental investigations of Baron-Ferejohn bargaining, the share of immediate agreements is very high, but some delay continues to occur also with experienced players. The percentage averages 76.1% and none of the pairwise mean comparisons between treatments is significantly different from zero. These patterns broadly support the corresponding implications of the Baron-Ferejohn model. We conclude that nominal asymmetries do not systematically impact on players' decisions regarding coalition size or their capacity to reach agreement.

Table 2. Coalition Size and Relative Coalition Weight by Treatment, Break and Experience

	A. Rel. Coalition Weight				**B. Coalition Size**			
	No Break		Break		No Break		Break	
	Periods				Periods			
	1-5	16-20	1-5	16-20	1-5	16-20	1-5	16-20
Base	0.74	0.72	0.75	0.68	3.70	3.58	3.76	3.42
	0.02	0.02	0.03	0.02	0.12	0.12	0.13	0.10
Pit1	0.75	0.67	0.74	0.71	3.68	3.43	3.62	3.48
	0.02	0.02	0.02	0.02	0.10	0.11	0.11	0.11
Pit2	0.70	0.65	0.76	0.76	3.50	3.36	3.74	3.84
	0.02	0.02	0.02	0.03	0.11	0.08	0.12	0.12

Notes. First row: means; second row: standard errors. Numbers are based on 50 final coalitions (46 for Pit1 with experience).

Result 5: *Contrary to Conjecture 3(b), we find that nominal weights do not systematically impact on relative coalition weight.*

4.4 Payoffs

Figure 6 provides players' average payoffs by weight-type, i.e., aggregated over their roles as proposers and responders. The left hand panels display an almost perfectly monotonic relationship between weight and payoffs for unexperienced players. In Pit1-nob (Pit1-b),

[19] Period-by-period probabilities of proposing a minimum winning coalition are displayed in Figure **??**. We also evaluated whether our results in Subsections 4.1 and 4.2 are sensitive to considering only proposals to minimum winning coalitions and found results to be qualitatively unchanged. See Table **??** for results on 'Proposer's Demand' as well as Figures **??** for Periods 1-5 vs. 16-20 and **??** for period-by-period results.

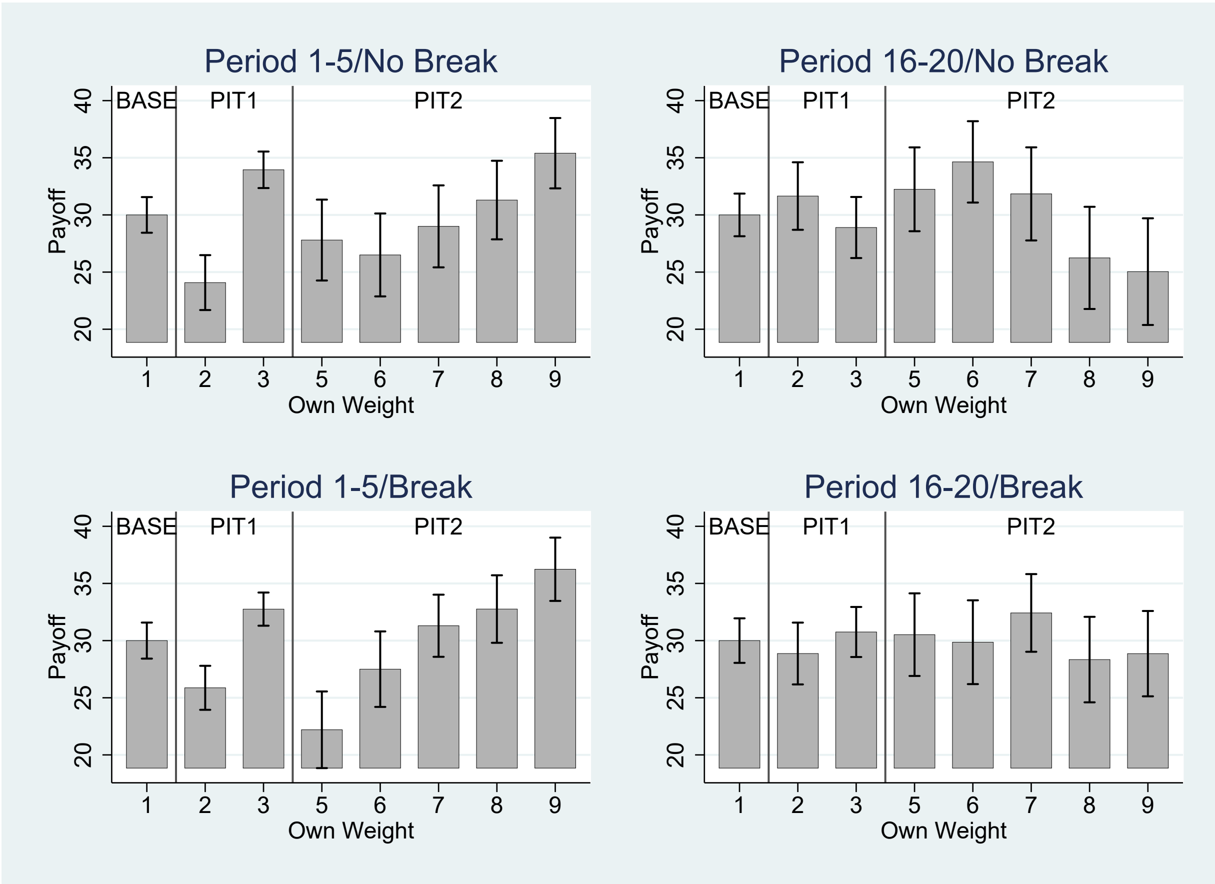

Fig. 6. Players' Payoffs by Weight, Break and Experience. The number of observations for each bar can be taken from the left panel of Table A6 in the Appendix. Error bars represent 90% confidence intervals.

unexperienced weight-3 players earned 9.87 (6.88) tokens more than unexperienced weight-2 players, corresponding to a gap of 41% (27%). Both player types' payoffs also differed significantly from weight-1 players in Base, where the average earning must always be 30: Weight-3 players earned 13% (9%) more and weight-2 players earning 20% (14%) less than weight-1 players.[20]

In Pit2-nob (Pit2-b), gaps between unexperienced weight-9 and weight-5 players amounted to 7.60 (14.04) tokens, or 27% (63%). Again both player types also differed significantly from weight-1 players in Base as weight-9 players earned 18% (21%) more and weight-5 players earning 7% (26%) less than weight-1 players. χ^2 tests reject equality of payoffs for unexperienced players in Pit2-nob ($\chi^2 = 7.95$, $p \leq 0.094$) and Pit2-b ($\chi^2 = 36.72$, $p \leq 0.01$).

The right panels of Figure 6 clearly indicate that experience had a massive impact on

[20]Mean, standard errors, case numbers and t-tests are given in the left panel of Table A6 in the Appendix. For the intermediate periods 6-10 and 11-15 please consult Figure B.12 in the Online Appendix.

payoffs. In Pit1 the payoff differential became insignificant, irrespective of a break. In Pit2, the effect of experience depended more strongly on whether there was a break or not. While the payoff difference between experienced weight-9 and weight-5 players turned negative (-7.20 tokens) in later periods without break, payoffs were not distinct from each other after a break. Overall, χ^2 tests cannot reject the null hypothesis of equality of payoffs for experienced players (Pit2-nob: $\chi^2 = 6.01$, $p = 0.199$; Pit2-b: $\chi^2 = 2.093$, $p = 0.719$).[21]

Result 6: *Unexperienced players' payoffs closely reflect nominal weight differentials in both* Pit1 *and* Pit2. *Experience levels out payoff differentials, except for* Pit2-nob, *where, as a consequence of discrimination against large-weight responders in later periods, we observe a tendency for 'strong' players to earn less than small weight players.*

5 Discussion

For unexperienced players, nominal weights often appear to provide focal points that they use in order to decide how much to allocate to themselves and to whom to make an offer. Conjecture 5(a) suggested that nominal weight differentials would cease to affect behavior and outcomes when players gain experience with the game. Yet, we find that experience does not *always* eliminate the influence of nominal weights, at least not over the course of 20 bargaining periods. The graphical impression from Section 4 suggests that outcomes were less biased across different weight types and came to resemble the corresponding baseline treatment more closely in treatments with break compared to uninterrupted play. Period-by-period results (see Figure B.11 and B.13 in the Online Appendix) underline this differential convergence. In this section, we first study the corresponding Conjecture 5(b) about the effect of the break. We then turn to the interaction of experience with the complexity of the bargaining situation (Conjecture 4).

5.1 The 'break effect'

A more detailed analysis revealed only few, and mostly unremarkable, *before-break* differences between treatments with and without break. The most notable one is that in treatments without break unexperienced weight-3 proposers claimed significantly more for themselves (t-tests, mean difference: 2.81, standard error: 1.10, $p = 0.011$). We observe the same tendency for unexperienced weight-9 proposers although the mean difference is insignificant here. Still it seems that even the announcement of a break has had some moderating impact on the demands of proposers endowed with the largest nominal voting power. Another difference pertained to responders' offers where the offers made to weight-5

[21] All results are also supported by GLS random effects panel regressions with session clustered standard errors and including subjects' sociodemographics (see right-hand panel of TableA6). The only difference worth mentioning is that, in Pit2-b, the difference between the payoffs of experienced weight-9 and weight-5 players (-1.85 tokens) became significant.

responders were significantly higher in treatments without a break (t-test, mean difference: 2.56, standard error: 1.36, $p = 0.061$). For both proposers' and responders' offers all other mean differences are negligible and insignificant. Pairwise comparisons further indicated that unexperienced 'strong' responders had an about 10% higher probability to receive an offer ≥30 tokens in Pit2-b compared to Pit2-nob.[22] Finally, we found only one significant difference in before-break payoffs, again in Pit2. Here, unexperienced weight-5 players underperformed in Pit2-b compared to Pit2-nob (mean difference: 5.6 tokens, standard error: 2.91, $p = 0.057$), which then gave rise to the very large payoff gap of more than 14 tokens relative to weight-9 players (cf. Subsection 4.4).

In contrast to these minor before-break differences, experienced play differed markedly between nob- and b-treatments. Table 3 quantifies the impact of the break for key outcomes, namely 'Proposer's Demand', 'Proposer's Offer (if> 0)' and 'Probability to Receive and Offer≥ 30', by means of a quasi difference-in-differences approach. The first three and the last three columns refer to Pit1 and Pit2, respectively. Consider 'Proposer's Demand'. We compute, for each subject, the difference between her demands when assigned the high-weight and the low-weight type for the first time (unexperienced) and when assigned the high-weight and the low-weight type for the last time (experienced). The cells of the table report the mean differences and the standard errors. The third and sixth columns labelled 'Difference Experience' report the within-subjects reduction of the bias due to experience. The row labelled 'Difference Break' reports the between-subjects difference of the bias between players in treatments without and with break. The difference of the differences, i.e., the impact of the break on the reduction of the bias due to experience, is given at the southeast position in boldface. Finally, we also report the impact of the break estimated by GLS regression with session clustered standard errors including subjects' sociodemographics. In the middle panel, we report corresponding data for 'Proposer's Offer (if>0)'. The bottom panel finally considers the impact of the break on differences in responders' probability to receive an offer≥30 tokens.

Before commenting on the results in Table 3, we note that the within-subject differences reported in the Table 3 cannot be directly compared to the differences between players with different weights reported in the previous section, which are averages over the first five and last five periods. Furthermore, since we concentrate here on individual subjects in two periods only, the number of observations per cell is distinctly lower ($n = 50$).

Focussing on 'Proposer's Demand', we find that the within-subjects bias in unexperienced players is very similar between nob and and b in both in Pit1 (at about 4.8 tokens) and in Pit2 (at more than 7 tokens). In Pit1, the bias is reduced significantly for experienced players only if treated with a break (it even slightly increases without break). The impact of the break on the reduction of power illusion amounts to more than 5 tokens. Although this result is insignificant, it is supported by a significant coefficient of the GLS regression (which is based on $n = 100$ observations). In Pit2, the break has

[22] The differences are (w-7: mean difference= 0.110, standard error= 0.045, $p = 0.015$; w-8: mean difference= 0.085, standard error= 0.041, $p = 0.038$; w-9: mean difference= 0.085, standard error= 0.036, $p = 0.018$; two-tailed t-tests).

less impact according to the raw data, and no impact at all when controlling for subjects' sociodemographics.

We observe a similar sized, yet insignificant, break effect (5 tokens) for 'Proposer's Offer (if > 30)' in Pit1. The results for Pit2 re-confirm Figure 2, where we observed that unexperienced proposers exhibit a bias against low-weight responders and experienced responders exhibit a bias against high-weight responders. The break, however, seems to dampen this effect in Pit2-b such that there is, at least according to the regression, a significant bias reduction of more than 6 tokens.

The bottom panel shows again that, in Pit1, the probability to get an offer$\geq$30 tokens is biased in favor of weight-3 responders (by more than 20 and 30 percentage points). Experience erases the bias and the break effect has the expected negative sign (albeit insignificant). In contrast, results for Pit2 impressively bring out the impact of the break on power-illusion bias: While unexperienced responders are exposed to the same bias of 38-40 percentage points, the bias only disappears for experienced players when treated with a break. Thus, the break effect amounts here to 40 percentage points. This result is also confirmed by regression, although the estimated coefficient turns out to be a bit lower. Summarizing, in treatments with a break, power illusion in experienced subjects was generally smaller or disappeared completely. Apparently, the break gives players a chance to reflect the strategic incentives of the bargaining situation more thoroughly. This 'break effect' is an interesting result by itself.

5.2 The effect of complexity

Given the strong similarity between the UWES treatment in Fréchette et al. (2005a) and our Pit1-b, it seems worthwhile to compare the two studies in some more detail. Main differences concern the size of bargaining groups (3 vs. 5 players), the nominal weighting schemes ($[50; 45, 45, 9]$ vs. $[7; 3, 3, 3, 2, 2]$) and the number of subjects (27 of whom 18 participated again as experienced subjects vs. 50 subjects per treatment). Fréchette et al. report larger requests by inexperienced proposers holding 45 votes compared to those holding 9 votes. Yet, no significant differences between the two voting weight-types exist in this or other respects for experienced players. For example, the null hypothesis that shares offered to weight-9 vs. weight-45 blocks are identical cannot be rejected. Similarly, in our experiment, weight-2 are initially distinct from weight-3 players regarding both their claims and how much they are offered by others and differences cease to be significant with players' experience and after the break. Despite the larger group size in our experiment, we conclude that results in the relevant treatment Pit1-b do not significantly differ from those reported by Fréchette et al. (2005a).

The Pit2 treatment introduced a greater complexity into the bargaining game. Conjecture 4 suggested that nominal asymmetries would thus shape players' expectations about others more strongly in Pit2 compared to Pit1. Indeed, we find some suggestive evidence that learning processes were slower in Pit2 treatments: When we focus on experienced subjects after the break in Tables A2 – A6, no statistically significant deviations between weight-3 and weight-2 players existed anymore in Pit1 (differences in Table A4 are small

Table 3. Impact of the Break on the Power-Illusion Bias

	PIT1 w-3 vs. w-2			PIT2 w-9 vs. w-5		
	Unexperienced	Experienced	Difference Experience	Unexperienced	Experienced	Difference Experience
			Proposer's Demand			
NOB	4.84*	6.98***	-2.14	7.44***	5.06***	2.38
	2.45	2.45	3.46	1.85	1.49	2.38
B	4.80***	1.36	3.44*	7.14***	3.22**	3.92
	1.28	1.70	2.14	2.42	1.29	2.74
Difference	0.04	5.62*	**-5.58**	0.30	1.84	**-1.54**
Break	2.76	2.98	4.17	3.05	1.97	3.55
Estimated			-5.05*			0.04
Difference			1.69			3.12
			Proposer's Offer (if>0			
NOB	3.80**	4.21**	-0.41	4.82	-12.82***	17.64***
	1.48	1.68	2.23	3.69	3.82	5.31
B	6.41***	1.79	4.63*	5.56*	-4.62	10.18**
	1.50	1.85	2.28	2.96	3.52	4.60
Difference	-2.61	2.42	**-5.04**	-0.74	8.20	**7.46**
Break	2.09	2.50	3.40	4.73	5.19	6.92
Estimated			-5.23			6.32*
Difference			3.16			2.64
			P(Offer≥30)			
NOB	21.7***	6.7	15.0	38.0***	-44.0***	82.0***
	6.6	9.9	11.9	8.0	9.0	12.1
B	32.7***	0.3	32.3***	40.0***	-2.0	42.0***
	7.5	9.3	11.9	7.6	10.1	12.6
Difference	-11.0	6.4	**-17.3**	-2.0	-42.0***	**40.0****
Break	10.0	13.6	17.5	11.0	13.6	17.1
Estimated			-25.3			29.4*
Difference			25.1			10.8

Notes. First row: Mean (mean difference) of the bias of offers (offer probabilities) towards the higher weight in tokens (percentage point); second row: standard errors. $n = 50$ in each cell. 'Difference Experience' and 'Difference Break' show the result of a two-tailed t test on the equality of two mean difference. The difference of the differences (impact of the break on power-illusion) is given in boldface. 'Estimated Difference' reports impact of the break estimated by GLS regression with session clustered standard errors including subjects' sociodemographics (Age, Gender, Number of Siblings, Labexperience, Field of Study, Works) as covariates ($n = 100$). Significance levels: $^{***}p \leq 0.01$, $^{**}p \leq 0.05$, $^{*}p \leq 0.1$.

and have the 'wrong' sign). For experienced weight-9 and weight-5 players, by contrast, differences that are statistically significant and in line with power illusion persisted with respect to three of the five outcome variables, namely in Tables A4 – A6.

6 Conclusion

We conducted an experiment designed to investigate whether and how "illusions" about voting power impact on legislative bargaining that follows closed rule Baron-Ferejohn procedures. To this end, we compared three scenarios where real bargaining power and the selection protocol were identical, but voting weights differed. The Baron-Ferejohn model predicts no difference in outcomes between the three scenarios. Further, we studied the effect of experience when the experiment proceeded with or without a break.

We find, first, evidence that differences in nominal voting power influence behavior and outcomes in multilateral bargaining for unexperienced players. This is especially true for proposers' claims, responders' probability to accept an offer and for proposers' decision on who should be made an offer. As a consequence, payoffs in the first bargaining periods reflected nominal weight asymmetries, and substantial unwarranted payoff heterogeneity existed even in the longer run (see upper-right panel in Figure 6). The experiment also revealed that experienced players are generally less swayed by nominal differences. Nonetheless, we found that, with uninterrupted play, some differences between nominally strong and weak players persisted, e.g., with respect to acceptance behavior. Other effects, notably the choice of coalition partners, were qualitatively altered by experience, but players continued to react to nominal weighting. By contrast, Fréchette et al. (2005a) report only transitory effects. We find that interrupting the experimental session with a break greatly expedited the transition towards non-illusionary behavior. Thus, the relevant findings in the two experiments are quite consistent with each other.

Bargaining processes in natural settings such as parliaments or international organisations regularly involve considerably larger committees than those studied in our experiment. Moreover, voting rules will usually be much less transparent because nominal and real power differentials will usually co-occur. Our experimental results suggest that potential 'power illusion' effects would be counterbalanced when negotiations are conducted by highly experienced and reflected agents. But, as a design recommendation, it seems safer to avoid any discrepancies between the distribution of voting weights or seats and the intended distribution of real power in the first place.

Given that our experimental design involved individual subjects switching between different voting weights, the extent of 'power illusion' in our data can plausibly be considered as a lower bound. There are reasons to conjecture that effects would be stronger if nominal weights are not randomly assigned but 'have a reason', e.g., being the representative of a larger population or a political party that has obtained more votes in a general election. Another interesting question is whether power illusion is contingent upon the bargaining protocol. Our experiment used the Baron-Ferejohn model because it is the leading legislative bargaining game and is widely used in political science. There is relatively little experimental work on other bargaining procedures, e.g., models of the demand bargaining type. These issues remain to be explored in future research.

Appendix A Additional Tables

Table A1. Breakdown of the Sample

	Mean	Std.Dev.
Age	25.0	4.4
Number of Siblings	1.6	1.1
	Freq.	Percent
Gender		
Female	183	61.0
Male	117	39.0
Labexperience[a]		
Yes	253	84.3
No	47	15.7
Works[b]		
Yes	206	68.7
No	94	31.3
Field of Study		
Economics	89	29.7
Other fields	51	17.0
Sciences	48	16.0
Social Sciences	31	10.3
Education Sciences	30	10.0
Humanities	22	7.3
Law	17	5.7
Philosophy	12	4.0

Notes. [a]Labexperience=Yes if participation in 3 or more experiments. [b]Student is employed.

Table A2. Proposer's Demand by Proposer's Own Weight, Break and Experience

		Raw Data				Predicted Values			
		Break				Break			
		No		Yes		No		Yes	
	Proposer's	Periods				Periods			
	Own Weight	1-5	16-20	1-5	16-20	1-5	16-20	1-5	16-20
BASE	1	41.92	47.23	42.87	45.05	41.91	47.22	42.41	44.58
		0.72	0.85	0.90	0.78	1.64	3.33	0.15	1.05
		250	250	250	250				
PIT1	2	34.99***	41.60***	34.53***	41.29**	35.11***	41.81*	35.03***	41.59**
		1.20	1.13	0.74	1.29	1.07	0.26	0.71	0.88
		100	92	100	92				
	3	42.50	46.36	39.69**	42.59**	42.63	46.49	40.03***	43.19
		0.86	1.21	0.68	0.77	0.66	0.26	0.73	2.35
		150	138	150	138				
	Diff.	7.51***	4.76***	5.16***	1.29	7.52***	4.68***	4.99***	1.61
	3 vs. 2	1.44	1.74	1.03	1.42	1.66	0.24	0.02	1.47
PIT2	5	38.20**	45.00	36.10***	40.80**	38.49*	43.97	36.96***	40.58**
		1.54	1.45	2.02	1.45	0.69	0.63	1.17	1.47
		50	50	50	50				
	9	46.74***	47.00	43.50	42.34**	45.32	47.68	43.64	42.84
		1.84	1.88	1.46	1.78	1.23	0.74	2.27	2.11
		50	50	50	50				
	Diff.	8.54***	2.00	7.40***	1.54	6.83***	3.71***	6.68***	2.26***
	9 vs. 5	2.40	2.38	2.49	2.30	0.55	0.12	1.11	0.64

Notes. Left panel: raw data. First row: mean demand; second row: standard error; third row: n. Right panel: predicted values (margins and contrasts) based on a GLS random effects panel model with session clustered standard errors including subjects' sociodemographics (Age, Gender, Number of Siblings, Labexperience, Field of Study, Works) as covariates. First row: mean; second row: standard error; $n = 2,960$ demands. Asterisks show the result of a two-tailed t test on the equality of the demands by proposers with own weights $\{2, 3, 5, 9\}$ and 1 (3 and 2, 9 and 5). Significance levels: $^{***}p \leq 0.01$, $^{**}p \leq 0.05$, $^{*}p \leq 0.1$.

References

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

Table A3. Proposer's Offer (if > 0) by Responder's Weight, Break and Experience

		Raw Data				Predicted Values			
		Break				Break			
		No		Yes		No		Yes	
	Proposer's	Periods				Periods			
	Own Weight	1-5	16-20	1-5	16-20	1-5	16-20	1-5	16-20
Base	1	32.21	34.08	32.86	37.92	33.27	35.52	32.96	38.66
		0.35	0.43	0.35	0.37	0.59	0.23	0.52	0.29
		839	754	815	692				
Pit1	2	25.94***	33.65	27.17***	33.73***	27.74***	35.45	28.09***	34.68
		0.57	0.73	0.52	0.71	1.23	0.65	0.16	3.62
		306	253	335	301				
	3	34.97***	35.99***	34.41***	34.60***	35.42***	37.33	34.98**	35.95
		0.41	0.63	0.37	0.56	0.57	1.76	0.81	1.76
		563	438	552	424				
	Diff.	9.03***	2.34***	7.24***	0.88	7.68***	1.89*	6.89***	1.27
	3 vs. 2	0.63	0.99	0.62	0.89	0.68	1.12	0.68	1.86
Pit2	5	30.82	38.78***	28.26***	37.66	31.60**	39.45***	30.02***	38.58
		1.07	0.79	0.82	0.79	0.58	0.70	0.14	0.63
		157	171	148	156				
	9	38.18***	35.99	37.07***	35.78**	39.27***	38.74	37.31**	37.89
		0.80	1.39	1.04	0.85	0.20	3.14	1.65	1.10
		169	113	187	135				
	Diff.	7.36***	-2.79*	8.81***	-1.88	7.67***	-0.71	7.29***	-0.70
	9 vs. 5	1.32	1.49	1.38	1.16	0.41	3.83	1.64	0.47

Notes. Left panel: raw data. First row: mean offer; second row: standard error; third row: n. Right panel: predicted values (margins and contrasts) based on a GLS random effects panel model with session clustered standard errors including subjects' sociodemographics (Age, Gender, Number of Siblings, Labexperience, Field of Study, Works) as covariates. First row: mean; second row: standard error; $n = 9,326$ offers. Asterisks show the result of a two-tailed t test on the equality of offers to respondents with weights $\{2, 3, 5, 9\}$ and 1 (3 and 2, 9 and 5). Significance levels: $^{***}p \leq 0.01$, $^{**}p \leq 0.05$, $^{*}p \leq 0.1$.

Ansolabehere, S., J. M. Snyder Jr., A. B. Strauss, and M. M. Ting (2005). Voting weights and formateur advantages in the formation of coalition governments. *American Jour-*

nal of Political Science 49(3), 550–563.

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

Baron, D. P., and J. A. Ferejohn (1989). Bargaining in legislatures. *American Political Science Review 83*(4), 1181–1206.

Berg, S., and M. J. Holler (1986). Randomized decision rules in voting games: A model for strict proportional power. *Quality and Quantity 20*(4), 419–429.

Bock, O., I. Baetge, and A. Nicklisch (2014). hroot: Hamburg registration and organization online tool. *European Economic Review 71*(C), 117–120.

Cutler, J., S. De Marchi, M. Gallop, F. M. Hollenbach, M. Laver, and M. Orlowski (2016). Cabinet formation and portfolio distribution in European multiparty systems. *British Journal of Political Science 46*(1), 31–43.

Diermeier, D., and R. Morton (2005). Proportionality versus perfectness: Experiments in majoritarian bargaining. In D. Austen-Smith and J. Duggan (eds.), *Social Choice and Strategic Decisions: Essays in Honor of Jeffrey S. Banks*, pp. 201–236. Berlin: Springer.

Diermeier, D., and S. Gailmard (2006). Self-interest, inequality, and entitlement in majoritarian decision-making. *Quarterly Journal of Political Science 1*(4), 327–350.

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

The Economist (2007, June 14th). The square root or death: The Germans fret that Poland may block a summit accord.
[Available at `http://www.economist.com/node/9341010/print`]

Eraslan, H. and A. McLennan (2013). Uniqueness of stationary equilibrium payoffs in coalitional bargaining. *Journal of Economic Theory 148*(6), 2195–2222.

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

Fréchette, G. (2009). Learning in a multilateral bargaining experiment. *Journal of Econometrics 153*(2), 183–195.

Fréchette, G., J. H. Kagel, and S. F. Lehrer (2003). Bargaining in legislatures: An experimental investigation of open versus closed amendment rules. *American Political Science Review 97*(2), 221–232.

Fréchette, G., J. H. Kagel, and M. Morelli (2005a). Nominal bargaining power, selection protocol, and discounting in legislative bargaining. *Journal of Public Economics 89*(8), 1497–1517.

Fréchette, G., J. H. Kagel, and M. Morelli (2005b). Gamson's Law versus noncooperative bargaining theory. *Games and Economic Behavior 51*(2), 365–390.

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

Fréchette, G., and E. I. Vespa (2017). The determinants of voting in multilateral bargaining games. *Journal of the Economic Science Association 3*(1), 26–43.

Freixas, J., and S. Kurz (2014). On minimum integer representations of weighted games. *Mathematical Social Sciences 67*(X), 29, 9–22.

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

Grimm, V., and F. Mengel (2011). Let me sleep on it: Delay reduces rejection rates in ultimatum games. *Economics Letters 111*(2), 113–115.

Kagel, J. H., H. Sung, and E. Winter (2005). Veto power in committees: An experimental study. *Experimental Economics 13*(2), 167–188.

Loewenstein, G. (1999). Experimental economics from the vantage-point of behavioural economics. *Economic Journal 109*(453), 25–34.

Mayer, A. (2018). Luxembourg in the Early Days of the EEC: Null Player or Not? *Games 9*(2), 29, 1–12.

McKelvey, R. D. (1991). An experimental test of a stochastic game model of committee bargaining. In T. R. Palfrey (ed.), *Contemporary Laboratory Research in Political Economy*, pp. 139–167. Ann Arbor: University of Michigan Press.

Miller, L., M. Montero, and C. Vanberg (2015). Legislative bargaining with heterogeneous disagreement values: Theory and experiments. CeDEx Discussion Paper Series No. 2015-24, University of Nottingham.

Oechssler, J., A. Roider, and P. Schmitz (2015). Cooling off in negotiations: Does it work? *Journal of Institutional and Theoretical Economics 171*(4), 565–588.

Miller, L., and C. Vanberg (2015). Group size and decision rules in legislative bargaining. *European Journal of Political Economy 37*(), 288–302.

Montero, M. (2006). Noncooperative foundations of the nucleolus in majority games. *Games and Economic Behavior 54*(2), 380–397.

Montero, M. (2017). Proportional payoffs in legislative bargaining with weighted voting: A characterization. *Quarterly Journal of Political Science 12*(3), 325–346.

Morelli, M. (1999). Demand competition and policy compromise in legislative bargaining. *American Political Science Review 93*(4), 809–820.

Nunnari, S., and J. Zapal (2016). Gambler's fallacy and imperfect best response in legislative bargaining. *Games and Economic Behavior 99*, 275–294.

Snyder, J., M. Ting, and S. Ansolabehere (2005). Legislative bargaining under weighted voting. *American Economic Review 95*(4), 981–1004.

Vespa, E. I. (2016). Malapportionment and multilateral bargaining: An experiment. *Journal of Public Economics 133*(1), 64–74.

Warwick, P.V., and J.N. Druckman (2006). The portfolio allocation paradox: An investigation into the nature of a very strong but puzzling relationship. *European Journal of Political Research 45*(4), 635–665.

Table A4. Responder's Probability of Voting 'Yes' as a Function of Responder's Share by Responder's Weight, Break and Experience: Contrasts

		No Break		Break	
	Responder's Share	Periods 1-5	16-20	1-5	16-20
	10	-.031**	-.021***	-.005***	.009***
		.013	.002	.001	.002
Pit1	20	-.089*	-.117***	-.053***	.047***
w-3		.052	013	.012	.006
vs.	30	-.123	-.239***	-.124***	.091***
w-2		.090	.043	.035	.010
	40	-.101	-.224***	-.094***	.074***
		.088	.059	.022	.011
	50	-.053	-.107***	-.029***	.026***
		.056	.040	.003	.004
χ^2	all	178.97***	219.44***	3,720.08***	2,074.20***
	10	-.002	-.044	.007	-.042***
		.005	.039	.009	.008
Pit2	20	-.029	-.156	.036	-.112***
w-9		.051	.127	.046	.024
vs.	30	-.080	-.264	.077	-.144***
w-5		.140	.200	.098	.015
	40	-.100	-.283	.085	-.113***
		.167	.180	.102	.009
	50	-.080	-.207*	.060	-.061***
		.111	.111	.065	.017
χ^2	all	5.37	219.44***	23.36***	4,297.66***

Notes. Contrasts based on a Probit random effects panel model with session clustered standard errors. Independent variable: Vote (No=0, Yes=1). Exogenous variables: Responder's Share (continuous) interacted with Responder's Weight, Break and Experience; Proposer's Share (continuous) interacted with Responder's Weight, Break and Experience; Responder's Share (continuous) interacted with Same Weight Dummy (only for Pit1). Regression includes subjects' sociodemographics (Age, Gender, Number of Siblings, Labexperience, Field of Study, Works) as covariates. First row: mean; second row: standard error. Pit1: $n = 1,000$ votes; Pit2: $= 425$ votes. Asterisks indicate significant contrasts; χ^2 value of jointly testing all contrasts. Significance levels: $^{***}p \leq 0.01$, $^{**}p \leq 0.05$, $^{*}p \leq 0.1$. The regression output is available from the authors on request.

Table A5. Responder's Probability of Getting an Offer$\geq$ 30 by Responder's Weight, Break and Experience

		Raw Data				Predicted Values			
		Break				Break			
		No		Yes		No		Yes	
	Responder's Weight	Periods 1-5	16-20	1-5	16-20	Periods 1-5	16-20	1-5	16-20
Base	1	0.67	0.63	0.66	0.68	0.70	0.67	0.71	0.73
		0.01	0.02	0.01	0.01	0.05	0.08	0.03	0.02
		1,000	1,000	1,000	1,000				
Pit1	2	0.41***	0.55***	0.46***	0.60***	0.42***	0.57	0.47***	0.62***
		0.02	0.03	0.02	0.03	0.01	0.08	0.02	0.02
		400	368	400	368				
	3	0.84***	0.68*	0.82***	0.67	0.87***	0.72	0.85***	0.69
		0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.08
		600	552	600	552				
	Diff.	0.42***	0.13***	0.36***	0.06**	0.45***	0.14*	0.38***	0.07
	3 vs. 2	0.03	0.03	0.03	0.03	0.01	0.08	0.01	0.10
Pit2	5	0.53***	0.80***	0.53***	0.76**	0.56**	0.82*	0.56**	0.81
		0.04	0.03	0.04	0.03	0.03	0.03	0.06	0.06
		200	200	200	200				
	9	0.81***	0.47***	0.89***	0.65	0.84**	0.50*	0.93***	0.71
		0.03	0.04	0.02	0.03	0.01	0.04	0.03	0.05
		200	200	200	200				
	Diff.	0.27***	-0.33***	0.36***	-0.11**	0.28***	-0.33***	0.36***	-0.11***
	9 vs. 5	0.05	0.05	0.04	0.05	0.02	0.01	0.04	0.02

Notes. Left panel: raw data. First row: mean probability; second row: standard error; third row: n. Right panel: predicted values (margins and contrasts) based on a Probit random effects panel model with session clustered standard errors including subjects' sociodemographics (Age, Gender, Number of Siblings, Labexperience, Field of Study, Works) as covariates. First row: mean; second row: standard error; $n = 11,840$ binary coded offers. Asterisks show the result of a two-tailed t test on the equality of the probability to get an offer>0 by responders with weight $\{2,3,5,9\}$ and 1 (2 and 3, 5 and 9). Significance levels: $^{***}p \leq 0.01$, $^{**}p \leq 0.05$, $^{*}p \leq 0.1$.

Table A6. Player's Payoff by Weight, Break and Experience

		Raw Data				Predicted Values			
		Break				Break			
		No		Yes		No		Yes	
	Player's	Periods				Periods			
	Weight	1-5	16-20	1-5	16-20	1-5	16-20	1-5	16-20
BASE	1	30.00	30.00	30.00	30.00	29.78	29.78	30.00	30.00
		0.94	1.13	0.96	1.18	0.30	0.30	0.21	0.21
		250	250	250	250				
PIT1	2	24.08***	31.65	25.87**	28.87	23.87***	31.41	26.07**	29.02
		1.45	1.78	1.16	1.63	0.42	1.10	1.58	0.96
		100	92	100	92				
	3	33.95***	28.90	32.75*	30.75	33.71***	28.70	32.96**	30.99*
		0.96	1.61	0.88	1.32	0.83	1.24	1.10	0.59
		150	138	150	138				
	Diff.	9.87***	-2.75	6.88***	1.88	9.85***	-2.71	6.89***	1.98
	3 vs. 2	1.67	2.54	1.43	2.09	1.24	2.33	2.67	1.53
PIT2	5	27.80	32.24	22.20***	30.52	27.56	32.74**	22.19**	30.75
		2.11	2.19	2.00	2.16	1.47	1.37	3.23	1.59
		50	50	50	50				
	9	35.40**	25.04*	36.24***	28.86	35.68***	25.03***	36.42**	28.90
		1.83	2.78	1.65	2.23	0.78	0.39	2.97	1.46
		50	50	50	50				
	Diff.	7.60***	-7.20**	14.04***	-1.66	8.12***	-7.72***	14.23**	-1.85***
	9 vs. 5	2.80	3.54	2.59	3.10	0.70	1.00	6.20	0.15

Notes. Left panel: raw data. First row: mean payoff; second row: standard error; third row: n. Right panel: predicted values (margins and contrasts) based on a GLS random effects panel model with session clustered standard errors including subjects' sociodemographics (Age, Gender, Number of Siblings, Labexperience, Field of Study, Works) as covariates. First row: mean; second row: standard error; $n = 2,960$ payoffs. Asterisks show the result of a two-tailed t test on the equality of the payoffs of players with own weights $\{2, 3, 5, 9\}$ and 1 (3 and 2, 9 and 5). Significance levels: $^{***}p \leq 0.01$, $^{**}p \leq 0.05$, $^{*}p \leq 0.1$.