

Polar Coalition Formation and Extreme Policies in Legislative Bargaining

A. Joseph Guse

Washington & Lee University

Abstract

I consider an n-player alternating offer bargaining game in which the allocation of private resources and a one-dimensional ideological good is decided. Importantly, players have heterogeneous preferences for the ideological good so that the composition of winning coalitions has real meaning. The main focus is characterizing the ideological composition of winning coalitions. I analyze a generalized ultimatum game as well as the infinite horizon game. I find that, in equilibrium, coalitions consisting of members running from one extreme up to the median can be expected under a wide variety of parameters. Moreover, such coalitions choose relatively extreme policies. I also show, contrary to previous results, that patience can be a force for more extreme policy choices.

Key Words: legislative bargaining; coalition formation; heterogeneous preferences

JEL Codes: C63, C78, D72.

INTRODUCTION

This paper brings together two stylized facts about legislatures. The first fact is that legislatures bargain over two very different sets of issues. On one hand, legislative bargaining redistributes wealth by deciding, for example, changes in tax policy or the funding of local public goods (a.k.a. “earmarks” and “pork”). On the other hand, legislatures set policies with ideological implications by deciding how to distribute funds across various (non-local) public goods (e.g. defense, scientific research, environmental protection, etc.) or how to regulate certain types of behavior (e.g. abortion, gun control, gay marriage, etc.). Real world legislation almost always contains a mix of ideological and redistributive implications. The second fact is that, to the extent that legislators can be positioned on an ideological spectrum, winning coalitions often form around one or the other of the extreme ends or poles of that spectrum. I call these *polar coalitions*.¹ For example, in the U.S. Congress, votes on substantive legislation tend to be along party lines with one party covering roughly the left half of the ideological spectrum and

¹ See Definition 1 for a precise description. “Polar” because they are defined by one or the other extreme polar ends of an ideological spectrum. In other words, if the task was to build a minimum majority contiguous coalition that had to include an extreme end, a polar coalition would result.

another party covering the right.² Conversely, center coalitions composed of moderates to the exclusion of extremists from both ends of the ideological spectrum are extremely rare. In this paper, I focus on the relationship between the bargaining environment in a legislative setting and ideological structure of winning coalitions. In particular, I ask whether the determination of ideological questions *alongside* the redistribution of private wealth contributes to the formation of polar coalitions.

The ideological make-up of winning coalitions is important because the policy enacted by a winning coalition reflects the ideological sentiments of the *coalition's* members. Also, since ideological decisions are public in nature, winning majorities impose their ideological will on the whole population. In this light, polar coalitions are of special interest. Compared to any other possible coalition structures they have the most extreme preferences. Therefore, if bargaining simultaneously over redistribution and ideological goods leads to polar coalition formation then legislatures are predisposed toward enacting relatively extreme ideological outcomes. Specifically, such outcomes would typically lie very distant from the median or the efficient policy choice depending on the distribution of preferences.

In order to understand the incentive to form polar coalitions and how robust it may be, this paper builds on Jackson and Moselle (2002) by using a model of alternating-offer majority-rule bargaining. The model has $n \geq 3$ players who decide on a one dimensional ideological good (represented by choosing a number between 0 and 1) and the allocation of a pot of money among the players' districts. All the players like getting as large a share of the pot for their district as possible, but have heterogeneous preferences for the ideological good. Preferences for the ideological good are summarized by ideal points on the $[0,1]$ interval. (Section 3 describes the formal model in detail.) As suggested above, there are at least two broad interpretations for the ideal points. One might imagine that the ideological question is how strictly to regulate an action such as abortion with 0.0 representing a policy of complete *laissez faire* and 1.0 representing a total ban. Alternatively, one might imagine that the ideological decision is how to allocate a second pot of money between two (non-local) public goods such as

² See Poole and Rosenthal (1991) for a detailed analysis of historical roll call patterns in the U.S. Congress.

environmental protection and defense. In this context, an ideal point of say 0.7 would represent a player who would most prefer that 70% of the second pot of money go to defense (*ceteris paribus* with respect to how the pot of money is distributed). I analyze two versions of this model - a generalized ultimatum game and an infinite horizon game.

One of the central insights generated by this model in both the ultimatum and infinite horizon versions is that the coalition choice problem can be thought of in terms of two competing forces – *conglomeration* and *polarization*. *Conglomeration* is an incentive to form close-knit formations, while *polarization* is an incentive to include extremists from both ends of the ideological spectrum. *Conglomeration* is fairly intuitive; members who are ideologically close make good coalition partners because the collective pain of compromising on the ideological good is low. *Polarization* is less intuitive; extremists are attractive as coalition partners because, in any round of bargaining, there will always be uncertainty over the outcome in the event that agreement cannot be reached. For extremists, the potential downside to disagreement is much worse than for moderates. This puts extremists in a weak bargaining position relative to moderates and lowers their demand for shares of the pot of money. It is this low demand for transfers of private wealth that makes them attractive.

Analysis of the model shows not only how these two forces – *conglomeration* and *polarization* – arise, but, more importantly how their *combination* leads to the formation of polar coalitions. I argue that legislative bargaining environments where these two forces exist in roughly equal measure are those most likely to sustain polar coalition formation. I highlight in particular how the discount factor affects the balance between *polarization* and *conglomeration*. Briefly, a higher discount factor (more patience) strengthens the *polarization* relative to the *conglomeration* effect, force making extremists more attractive to coalition builders. This is because the weak bargaining position of extremists derives from uncertainty over how bargaining might play out in later rounds. If nobody cares about the future, the basis for this difference between extremists and moderates evaporates and only *conglomeration* would matter.

For the ultimatum game, I derive a sufficient condition for polar coalitions to form the basis of any subgame perfect equilibrium. Using that sufficient condition, I show that under a squared distance loss function, a discount factor equal to one and equal

expected private wealth values, polar coalitions must form in any subgame perfect equilibrium for *any* distribution of ideal points. This provides a useful benchmark for arguing that polar coalitions should frequently arise in equilibrium when conglomeration and polarization are sufficiently balanced.

For the infinite horizon game, I derive a necessary and sufficient condition for polar coalition formation in a stationary equilibrium when the distribution of ideal points is symmetric. An examination of this condition demonstrates that the roles played by conglomeration and polarization and how they are balanced by the discount factor are robust to changes in assumptions made about the length of the bargaining horizon. I also search numerically for polar-coalition-based stationary equilibria in the infinite horizon game. This analysis confirms the broad appeal of polar coalitions across assumptions on the functional form of players' utility functions, the discount factor, the number of players in the game and the distribution of players' ideal points. Furthermore, the numerical results further illustrate the role of the discount factor, suggested by the theory, in balancing the conglomeration and polarization incentives.

The paper proceeds as follows. Section 2 describes related literature. Section 3 introduces the basic model. Section 4 presents the results for the generalized ultimatum game. Section 5 presents the theoretical and numerical analysis of the infinite horizon game. Section 6 concludes.

LITERATURE REVIEW


Baron and Ferejohn (1989) extended the two-player alternating-offer divide-the-dollar bargaining game famously analyzed in Rubinstein (1982) by considering $n > 2$ players. Jackson and Moselle (2002), upon which my analysis builds, added a public ideological good to the model. They proved the existence of "simple" equilibria which are characterized by stationarity, immediate agreement and minimum majority coalitions. They did not, however, put much focus on the composition of coalitions except to say that every player has some chance of being excluded in every round. By contrast, coalition composition is the central question here.

Apart from the model of Jackson and Moselle (2002), I am not aware of any other models in the tradition of Baron and Ferejohn (1989) that combine a public or ideological good and heterogeneous preferences for it.³ As such, most Baron-Ferejohn extensions or derivatives cannot speak to the main issue addressed in this paper: the ideological composition of winning coalitions. Examples of Baron-Ferejohn extensions with heterogeneity in things other than preferences for a public or ideological good include Calvert and Dietz (1996) who admit externalities by allowing players to care about other players' share, Harrington (1989) who allows heterogeneous preferences for risk and Norman (2002) who looks at the heterogeneity of time preferences.⁴ Lizzeri and Persico (2001), Volden and Wiseman (2007), Battaglini and Coate (2007, 2008) are all examples of models where bargaining takes place over both private wealth distribution and a public good, but where preferences for the public good are homogeneous. Such models are well suited for investigating the overall provision of public goods, but cannot speak to the composition of coalitions.

Other political bargaining models have players with heterogeneous preferences over a public or ideological good space. In many of these, the question of coalition structure does arise. Unfortunately, many of these limit their analysis to the case of three players (Austen-Smith & Banks, 1988; D. Baron, 1991; D. P. Baron & Diermeier, 2001; D. P. Baron, Diermeier, & Fong, 2007). Since three players can only form three majority coalitions⁵, it is difficult to make many general claims about coalition composition which might extend to larger games. Banks and Duggan (2000, 2006) and Baron (1996), discussed below, are notable exceptions allowing any number of players.

³ Banks and Duggan (2000,2006) is very general and could, in principle, accommodate a game where players bargain over both an ideological good and private wealth, but they do not specifically analyze coalition formation or policy choice for that case.

⁴ Norman (2002) finds a link between the heterogeneity of time preferences and uniqueness of subgame perfect equilibria and the possibility of non-minimum majority coalition formation. This possibility in the standard Baron-Ferejohn model is due to the symmetry of homogeneous time preferences and proposer indifference over who to include in their coalitions. He shows how the introduction of heterogeneous time preferences destroys this symmetry, making subgame perfect behavior in each round determinant and unique. In this paper, heterogeneous preferences for a public good instead of time preferences destroys that symmetry ruling out multiplicity of equilibria and super-majority winning coalitions in the finite horizon game.

⁵ 

The literature on electoral competition, as opposed to bargaining, has long discussed exceptions to the median voter hypothesis. For example, Palfrey (1984) shows how the threat of a third party can give incentive to two dominant parties to take significantly divergent positions. Ingberman and Villani (1993) construct a model in which parties run for an executive position and legislative seats. They show, under a variety of specifications of voter behavior, how the interaction between the legislative and executive branches in the policy setting process can give incentive for parties to take divergent positions in the election stage. Alesina and Rosenthal (2000) examine a model of elections similar to that in Ingberman and Villani in which voters again elect both an executive and a legislature. They show how split ticket voting gives parties an incentive to take extreme platform positions. The main conclusions in this paper complements these results by demonstrating how outcomes representing significant deviations from the median voter can arise in a bargaining game.

Polar coalitions, as I define them, also appear in Baron (1996) and the one dimensional case analyzed in Banks and Duggan (2006).⁶ However, I arrive at very different conclusions about the mechanisms through which they arise and how they behave. In contrast to both of these models, policies formed by polar coalitions in my model do not converge to the median. Furthermore, Banks and Duggan (2006) show that a high discount factor strengthens the draw toward the median policy. They summarize this result by saying that “legislative patience implies policy moderation” suggesting that this relationship generalizes beyond the one-dimensional case generating the result.⁷ While the relationship between the discount factor and equilibrium policy choice is more subtle in my model, it is nearly opposite in spirit. Specifically, I show that proposers accommodate the policy preferences of their coalition members and that patience encourages proposers to include extremists in their coalitions. The apparent contradiction can be explained by the lack of private wealth

⁶ Besides these two cases, one might interpret the coalitions in Romer and Rothenthal (1979) as polar coalitions. In that model, however, the proposer’s ideal point is assumed to be an extreme point. Therefore, it is not surprising that the winning coalition is the proposer’s half of the ideological spectrum

⁷. See especially Banks and Duggan (2006, p.52 and p.60)

transfers in both Baron (1996) and Banks and Duggan (2006).⁸ Without cash transfers, there is no mechanism to extract coalition members' surplus utility and thus no incentive to choose the policy which maximizes the coalition's utility. This leads to policy choices which tend toward the median ideal point. As such, the distinction between the right half and left half as winning coalitions is nearly payoff irrelevant. By contrast, in the model employed here, policy choices maximize coalition utility in equilibrium. This translates to policies located toward the center of the *coalition's* ideal points⁹ as opposed to the center of the *population's* ideal points. Therefore, right and left polar coalitions choose very different and potentially extreme policies (depending on the distribution of ideal points) with significant payoff consequences.

Another strand of the political science literature explores the question of party discipline in legislatures. Prime examples include Krehbiel (2000) and McCarty et al (2001). According to this view of legislative bargaining, political parties are exogenous entities representing the left and right halves of the political spectrum that enforce voting blocks similar to the polar coalitions described in this paper. However, in contrast to framework employed in that literature, the model in this paper makes no assumptions about political parties. Coalitions representing the left and right half of the ideological spectrum arise endogenously out of the basic elements of the bargaining game.

Hence, the novelty in this paper is not the model itself. Nor is this the first paper to describe or attempt to explain polar coalitions as I define them here. Instead, this paper offers a new story about why such coalitions form and their relatively extreme policy choices.

⁸ Banks and Duggan (2000,2006) employ a very general model which could, in principle, accommodate a game where players bargain over both an ideological good and private wealth, but they do not specifically analyze coalition formation or policy choice for that case.

⁹ For example, if the ideological loss function takes the typical squared-distance form, then the policy is exactly the coalition mean ideal point. If the loss function is the absolute distance, it is the median. This is a consequence of Proposition 2, part (ii), below.

MODEL

There is a set of $n \geq 3$ players denoted $N = \{0, 1, \dots, n-1\}$.¹⁰ Each player has preferences defined on a set of *proposable outcomes*, $\Theta \subset \{\mathbb{R}^n \times [0, 1]\}$. A proposal $\theta \in \Theta$ consists of two items – a policy $x(\theta) \in [0, 1]$ and an allocation $\{y_j(\theta)\}_{j \in N} \in \mathcal{Y} \subset \mathbb{R}^n$ of n dollars among the n members so that $\mathcal{Y} = \{y \in \mathbb{R}^n, \sum_{j \in N} y_j \leq n, y_j \geq 0\}$. Each player's preferences are common knowledge and are characterized by an *ideal point*: Let $x_j \in [0, 1]$ denote the policy individual j would most like to see enacted. Players' labels are in order of ideal point, so that player 0 has the lowest ideal point and player $n-1$ has the highest.

I consider two versions of the game – a generalized ultimatum game and an infinite horizon game. In both versions, each player is selected with probability $\frac{1}{n}$ to be the *proposer* in every round of bargaining $t \in \{0, 1, \dots\}$. The proposer announces some proposal $\theta^t \in \Theta$. After learning the proposal, each player, including the proposer, votes either YES or NO. Players vote in a fixed sequence.¹¹ If at least m players vote YES, the proposal is enacted and the game ends (in both versions). In the ultimatum game, if fewer than m vote YES, a default outcome, θ^1 , drawn from a commonly known distribution, is implemented. In the infinite horizon game, the game advances to the next round of bargaining.

For a given outcome $\theta \in \Theta$, player j cares about her own share of the cash and how far the policy is set from her own ideal point according to the (undiscounted) payoff function $u_j: \Theta \rightarrow \mathbb{R}$.

¹⁰ In order to keep the notation consistent, small-case roman letters are used for numbers - either as quantities or indices, while capital roman letters denote sets or quantities aggregated over sets. Capital script letters (e.g. \mathcal{A}, \mathcal{M}) denote collections of sets.

¹¹ The order of voting itself is not important. Simultaneous voting could be accommodated as well, but with some difficulty, since it would admit subgame perfect equilibria in which majority interests do not translate into majority voting. For example, even if everyone was in favor of a proposal, everyone might vote no in a simultaneous voting game if they believed everyone else would. Proposition 1 says that this cannot happen with sequential voting. Authors working with similar models take a variety of approaches to the same effect. For example, like this paper, Jackson and Moselle (2002) assume sequential voting, while Banks and Duggan (2006) impose a restriction on voting behavior called "weak dominance" which simply means that players vote YES if and only they are weakly in favor.

$$u_j(\theta) = y_j(\theta) - g(|x(\theta) - x_j|) \quad (1)$$

g represents the ideological loss felt by player j with ideal point x_j from having public policy $x(\theta)$ implemented. I assume that g is twice continuously differentiable with $g(0) = 0$, $g(1) = 1$, $g' \geq 0$ and $g'' > 0$. If agreement on θ is reached in round t , j 's payoff for the game is $\delta^t u_j(\theta)$ with $\delta \in [0, 1]$.

Whenever expectations are well-defined, $E(z^t)$ denotes the expected value of some round t random variable z just before the identity of the round t proposer is revealed. In particular, $E(u_j^t)$ denotes j 's expected payoff at the beginning of round t . When discussing players' strategies in round t , j 's *continuation value* is $\delta E(u_j^{t+1})$. If player j prefers a proposal, θ^t , in round t over continuation so that $u_j(\theta^t) \geq \delta E(u_j^{t+1})$, we say that player j *favors* θ^t or *strictly favors* if the inequality holds strictly. $C(\theta^t) \subset N$ denotes the set of players or *coalition* that favors θ^t .

Subgame perfection is the solution concept. Since in each round, players propose and vote in sequence and observe all previous actions, this is a game of perfect information and subgame perfection has its usual interpretation: Players' strategies are subgame perfect, if behavior (either proposing or voting) at every decision node leads to a distribution of outcomes yielding the highest expected payoff for the player moving at that node, whether or not that node is reached in equilibrium (Selten, 1975).¹²

Strictly speaking, a player's strategy must describe a proposal to announce in the event that player is selected as the proposer as well as a description of voting behavior for any given history of play. Proposition 1, however, allows us to ignore the details of players' voting behavior.

Proposition 1. *In any subgame perfect equilibrium, if m players strictly favor a proposal, then it would pass with probability one. Similarly, if m players strictly prefer continuation to the proposal, then it would fail with probability one (Proof: see Appendix A).*

¹² If the terminal outcome is uncertain, it is still drawn from a known distribution so that continuation values in the last round of bargaining are well defined.

In other words, Proposition 1 says that while the model does not require voters to vote their interest, the voting outcomes, in terms of whether a given proposal passes, is always identical to an environment where players are assumed to vote their interests.

THE GENERALIZED ULTIMATUM GAME

Jackson and Moselle (2002) showed that in the infinite horizon game, there always exist equilibria in which agreement is reached in every round of bargaining. Rather than redo their analysis, the goal here is to focus on the coalition choice problem in the first round of bargaining and the question of coalition composition. Toward that end, I analyze a truncated game. This is equivalent to an ultimatum game in which disagreement in the first round results in whatever outcome, θ^1 , would have been agreed upon in the second round ($t = 1$). In the most general truncated game, this outcome consists of $x(\theta^1) \in [0, 1]$ and an allocation $\{y_i(\theta^1)\}_{i \in N}$. In particular, commonly held beliefs about the distribution of θ^1 implies *some* vector of expected cash shares, $\{E(y_i(\theta^1))\}_{i \in N}$ and *some* distribution of policy values, $F(x(\theta^1))$. It follows from (1) that undiscounted continuation values for all $i \in N$ are given by the following.

$$E(w_i(\theta^1)) = E(y_i(\theta^1)) - \int_0^1 g(|x_i - x(\theta^1)|) dF(x(\theta^1)) \quad (2)$$

Such a truncated game is very general. For example, any simple equilibrium in the infinite horizon game of Jackson and Moselle (2002) would be a special case. Unfortunately, it is a bit too general to be tractable¹³ and I restrict attention to games where n is odd, $m = \frac{n+1}{2}$, F puts weight on at least two policy outcomes, and the expected cash shares are equal. Formally,

$$E(y_i(\theta^1)) = 1 \forall i \quad (3)$$

The last assumption, which is dropped for the analysis in the next section, focuses attention on how the heterogeneity of ideal points - as opposed to private wealth

¹³ In particular, it can be shown that the most general truncated game leaves open the possibility that continuation values for some players be negative. For example, just construct a default outcome where one player gets all the cash with probability one. However, the result that winning coalitions are minimum majority (see Proposition 2) depends on all players having positive continuation values.

expectations – shape coalition formation. Proposition 2 states that the proposer makes an immediately passable proposal with the support of a minimum majority coalition. This proposal consists of a policy which minimizes ideological loss for the coalition and cash shares which make non-proposing coalition members just indifferent between the proposal and continuation leading to the formulation for i 's payoff in equation (6).

Proposition 2. *In a subgame perfect equilibrium of the truncated game, the proposal θ^0 made by player i is characterized by the following.*

- i. θ^0 will pass and the game will end with $i \in C(\theta^0)$.
- ii. $\#C(\theta^0) = m$ and for each $j \in C(\theta^0)$,

$$y_j(\theta^0) = \partial B u_j(\theta^1) + \rho (|x_j - x(\theta^0)|) > 0 \quad (4)$$

Of the all the proposals favored by $C(\theta^0)$, θ^0 uniquely maximizes $u_i(\theta)$. In particular, $x(\theta^0)$ uniquely minimizes $C(\theta^0)$'s ideological loss (Proof: see Appendix).

$$x(\theta^0) = \underset{x}{\operatorname{argmin}} \sum_{j \in C(\theta^0)} \rho (|x - x_j|) \quad (5)$$

i 's payoff is given by (Proof: see Appendix).

- iii. i 's payoff is given by (Proof: see Appendix).
- iv.

$$u_i(\theta^0) = n + \partial B(u_i^1) - \rho \sum_{j \in C(\theta^0)} B(u_j^1) - \sum_{j \in C(\theta^0)} \rho (|x_j - x(\theta^0)|) \quad (6)$$

COALITION CHOICE, CONGLOMERATION AND POLARIZATION

Proposition 2 parallels or synthesizes points already made by Jackson and Moselle (2002) for the specific set-up used here. Nevertheless, it is worth explaining how it sets the stage for the coalition choice problem which, again, is the emphasis here. Toward that end, note that (iii) essentially implies that the policy choice, $x(\theta^0)$ is the average of the coalition members' ideal points. Furthermore, the uniqueness of $x(\theta^0)$ for a given coalition and therefore the uniqueness of θ^0 means that $C(\theta^0)$ (with appropriately

restricted domain) is bijective implying an inverse map, $\theta^0(C)$. One can now express everything in terms of the coalition choice, C , instead of the proposal, θ^0 . In particular, I use $u_i(C)$ and $x(C)$ as shorthand for $u_i(\theta^0(C))$ and $x(\theta^0(C))$.

It follows from Proposition 2 that the problem faced by player i in the role of proposer is to choose a coalition C of size m to which i belongs that maximizes $u_i(C)$. The proposer's payoff from (6) can be decomposed as follows.

$$u_i(C) = n + \delta B(u_i^1) - \delta U^1(C) - G(C) \quad (7)$$

where

$$G(C) = \sum_{j \in C} g(|x_j - x(C)|)$$

$$U^1(C) \equiv \sum_{j \in C} B(u_j^1) = \sum_{j \in C} \left[1 - \int_0^1 g(|x_j - x^1|) dF(x^1) \right]$$

The first two terms in (7) do not depend on the choice of coalition, C . Therefore, maximizing $u_i(C)$ is exactly equivalent to minimizing $G(C)$ and $U^1(C)$ whose sum I refer to as C 's *transfer demand*. *Conglomeration* is the incentive to minimize $G(C)$ which decreases the more tightly packed C 's ideal points are. *Polarization* is the incentive to minimize $U^1(C)$ which decreases the more extreme C 's ideal points are.

In order to understand how these incentives guide coalition choice, recast the discrete coalition choice problem as a continuous choice. Imagine that the proposer can adjust up or down the ideal points of his coalition members. If C is a coalition of size m containing player a with fixed ideal point x_a , let z_a be a hypothetical player with variable ideal point, x_{z_a} . Let $C_{z_a} \equiv \{C \setminus a \cup z_a\}$ be the coalition where a is replaced by z_a . The approach is to consider the affect of changes in x_{z_a} on the proposer's payoff through $G(C_{z_a})$ and $U^1(C_{z_a})$ as well as $G_f(C_{z_a})$ defined as follows.

$$G_f(C_{z_a}) = \sum_{j \in C_{z_a}} g(|x_j - x(C)|) \quad (8)$$

G_F is the ideological loss members of C_{z_a} would experience under a policy $x(C)$ which is the loss minimizing choice for coalition C (but not necessarily for C_{z_a}). I refer to the derivative of $G_F(C_{z_a})$ - where we hold the policy fixed at $x(C)$ even as we vary the x_{z_a} - as the *field conglomeration effect*. The *individual conglomeration effect* is the derivative of $G(C_{z_a})$ where we allow $x(C_{z_a})$ to adjust. The *polarization effect* is the derivative of $U^1(C_{z_a})$

Figure 1 illustrates the key features of the conglomeration effect. The upper panel shows the totals $G(C_{z_a})$ and $G_F(C_{z_a})$ (in bold), while the lower panel shows the first derivatives of these objects. The main thing to note is that $G_F(C)$ reaches its minimum at the center of C 's ideal points (as defined by $x(C)$). The "effect" (the derivative of G_F) is negative to the left of $x(C)$ because increasing a 's ideal point from a position in that region would reduce $G_F(C_{z_a})$ (which is beneficial to the proposer). On the other hand, it is positive to the right of $x(C)$ since increasing a 's ideal point from a position in that region would increase $G_F(C_{z_a})$ (which hurts the proposer). Overall, the conglomeration effect represents an incentive to move ideal points closer to $x(C)$. The figure also illustrates the subtle difference between $G_F(C_{z_a})$ and $G(C_{z_a})$. Note, in the lower panel, that the field effect is everywhere steeper than the individual effect and that the individual and field curves intersect at x_{z_a} , the ideal point of the player replaced by the hypothetical player z_a . The convenience of working with G_F can be seen by noting that while in general, for two distinct players $a \in C$ and $b \in C$, $G_F(C_{z_a}) \neq G_F(C_{z_b})$, they are vertical shifts of each other so that the graphs of $\frac{d}{dx_{z_a}} [G_F(C_{z_a})]$ and $\frac{d}{dx_{z_b}} [G_F(C_{z_b})]$ are identical. Therefore, when speaking of the marginal field effect, it does not matter which member of coalition C the hypothetical z player replaces.¹⁴ This also explains why the derivative of $G(C_{z_a})$ is called the *individual effect*; it depends on the choice a , while the derivative of $G_F(C_{z_a})$ does not (Figure 1).

¹⁴ $G_F(C_{z_a})$ and $G(C_{z_a})$ are characterized formally in the proof of Proposition 4 in the appendix. See especially equations (19) through (25).

Figure 2 illustrates $U^1(C_{z_a})$ and its derivative, the polarization effect. Define \hat{x} as the point where $U^1(C_{z_a})$ reaches its maximum or where its derivative is zero. It is the hypothetical ideal point of the player with the lowest expected ideological loss (highest expected payoff) in the default outcome. Since the proposer must compensate coalition members who expect higher payoffs in the default outcome with greater shares of the pot of money, \hat{x} is the worst choice for the proposer (holding C constant). Hence the polarization effect represents an incentive to move ideal points away from \hat{x} . In the upper panel of Figure 2, we see two versions of U^1 – one for some coalition C whose member a has been replaced by z_a and another coalition, C' whose member b has been replaced by z_b . In the lower panel, however, we see that the derivatives of these two functions are identical. ($\frac{d}{dx_{z_a}} U^1(C_{z_a})$ is constant with respect to the choice of C or a .) Therefore, there is only one polarization effect and only one \hat{x} no matter which coalition we are talking about (Figure 2).¹⁵

Proposer i 's polar coalitions, L_i and R_i , are formally defined as the left and right-most coalitions of size m to which i belongs.

Definition 1. For any player $i \in N$, define L_i and R_i as follows.

$$L_i = \begin{cases} \{0\}_{j=0}^{m-1}, & i \leq m-1 \\ \{0\}_{j=0}^{m-2} \cup i, & i > m-1 \end{cases} \quad (9)$$

$$R_i = \begin{cases} \{0\}_{j=n-m+1}^{n-1} \cup i, & i \leq n-m \\ \{0\}_{j=n-m}^{n-1}, & i > n-m \end{cases} \quad (10)$$

Proposition 3 says that the policy choice has a positive relationship with the position of the coalition members' ideal points. It follows from point (0) of Proposition 2 and it has an immediate implication: policy choices are more extreme under polar coalitions than

¹⁵ $U^1(C_{z_a})$ is characterized formally in the proof of Proposition 4 in the appendix. See in particular equations (26) and (27).

under any other minimum majority coalitions a proposer might form. Proposition 4 is the central result of this section; it provides a sufficient condition under which polar coalitions form in equilibrium.

Proposition 3. For all C and any $\alpha \in C$, $0 < \frac{d}{dx_{\alpha}} x(C) < 1$ (Proof: see Appendix).

Proposition 4. In the truncated game for any set of ideal points, any default policy distribution $F(x^1)$, and equal expected cash shares, if for all C and any $\alpha \in C$, we have

$$(9) \quad \text{sign} \left(\frac{d}{dx_{\alpha}} [G_f(C_{\alpha}) + \delta U^1(C_{\alpha})] \right) = \text{sign}(\bar{x} - x(C)) \quad (11)$$

for all $x_{\alpha} \in [0, 1]$, then proposer i will always choose to form either L_i or R_i in the subgame perfect equilibrium (Proof: see Appendix).

Recall that the proposer's problem is to choose a coalition C that minimizes the sum of $G_f(C) + U^1(C)$, but we are recasting the problem as one repositioning the ideal point of some member z_{α} in order to minimize $G_f(C_{\alpha}) + U^1(C_{\alpha})$. The sum of the polarization and conglomeration effects is the sum of the derivatives of these functions and, as such, represents the rate of change in transfer demand as the proposer "moves" the z_{α} 's ideal point to the right. Proposition 4 says that if this rate is negative for all coalitions which are already right-leaning ($\bar{x} - x(C) < 0$), then R_i must be the best choice among right-leaning coalitions. Similarly, if the sign is positive for all left-leaning coalitions ($\bar{x} - x(C) > 0$), L_i must be the best choice among left-leaning coalition. The proof, provided in the appendix, makes this argument precise and also explains why the condition in Proposition 4 uses the field conglomeration effect (which is independent of the choice of α) in place of the individual conglomeration effect.

Figure 3 illustrates the intuition behind Proposition 4 for the case of a right-leaning coalition – one where $\bar{x} - x(C) < 0$. The proposer is always rewarded with a lower value of $G_f(C_{\alpha})$ whenever z_{α} 's ideal point gets pushed toward $x(C)$ and is rewarded with a lower value of $U^1(C_{\alpha})$ whenever z_{α} 's ideal point gets pushed away from \bar{x} . With this in mind, note that Figure 3 divides the interval into three regions. In region II, pushing an

ideal point to the right accomplishes both goals; increasing x_{α} 's ideal point move it toward $x(C)$ and away from \bar{x} . However, outside of II, there is always a trade-off. In Region I, for example, pushing an ideal point to the right would increase $U^1(C_{\alpha})$ but would decrease $G_F(C_{\alpha})$. In other words, such a move would work against the polarization effect but would be rewarded by conglomeration effect. Figure 3 illustrates a particular case where the ambiguity in Region I is resolved by the fact that conglomeration effect dominates the polarization effect in that region. Furthermore, in Region III, the opposite hold true. Therefore condition (11) holds over the entire interval for the coalition, C , under inspection. Proposition 4 put another way says that if a picture like Figure 3 holds for all right-leaning coalitions, then the polar right coalition is the best choice among all right-leaning coalitions (Figure 3).

Keep in mind that Proposition 4 does not guarantee that (11) will always hold; it simply says that (11) is a sufficient condition for polar coalition formation. In Section 4.3 (below), I argue that while (11) can indeed be violated, there is good reason to believe that it should hold under a broad set of circumstances. Before making that case, it is helpful to consider a benchmark case for which (11) is guaranteed to hold.

4.2. Example: Squared Loss. Suppose that ideological loss was equal to the square of the distance from the policy x , so that

$$g(|x_j - x|) = (x_j - x)^2 \quad (12)$$

In this case, it is trivial to show that the polarization and field conglomeration effects are given by the following.

$$\begin{aligned} \frac{d}{dx_{\alpha}} [U^1(C_{\alpha})] &= -2(x_{\alpha} - \bar{x}) \\ \frac{d}{dx_{\alpha}} [G_F(C_{\alpha})] &= -2(x_{\alpha} - x(C)) \end{aligned}$$

where, due to the squared loss assumption (12), \bar{x} is simply equal to $\int_0^1 x^2 dF(x^2)$, the expected value of the default policy and $x(C)$ is the mean of C 's ideal points. Plugging

these into (11) and gathering terms, checking the sufficient condition for some coalition C comes down to asking whether

$$\text{sgn}\left((1 - \delta)x_{z_a} + \delta\hat{x} - x(C)\right) = \text{sgn}(\hat{x} - x(C)) \quad (13)$$

First note that when $\delta = 1$, this condition holds for all values of x_{z_a} and all coalitions C . In other words, when players put full weight on the continuation outcome, subgame perfection requires polar coalition formation *no matter* the distribution of players' ideal points and *no matter* beliefs about the policy outcome in the continuation game.¹⁶ Second, if $\delta < 1$, (13) is more likely to be violated the smaller the distance between \hat{x} and $x(C)$ and the closer x_{z_a} is to either 0 or 1. This is because a lower value of δ weakens the polarization effect relative to the conglomeration effect.

4.3. Robustness Part 1: Curvature of g .

What if the loss function g is not quadratic? To fix ideas, focus on the case where $x(C) > \hat{x}$ as shown in Figure 3. Note that in region II, between \hat{x} and $x(C)$, both components of transfer demand, $V^1(C_{z_a})$ and $G_j(C_{z_a})$, are decreasing in z_a 's ideal point giving the proposer an unambiguous incentive to increase the value of z_a 's ideal point. Outside of that interval, there is ambiguity. (See arrows in Figure 3.) In region III, to the right of $x(C)$, the polarization effect recommends increasing x_{z_a} , while the conglomeration effect recommends decreasing. In region I, to the left of \hat{x} , the opposite is true. Therefore, if there are violations of condition (11), it is either because conglomeration fails to dominate in region I or because polarization fails to dominate in region III. Moreover, it is clear from Figure 3 that (11) must survive outside of region II in at least a small neighborhood left of \hat{x} and right of $x(C)$. To see why, note that at \hat{x} , $\frac{d}{dx_{z_a}}[V^1(C_{z_a})] = 0$ while $\frac{dG_j(C_{z_a})}{dx_{z_a}}$ is strictly negative and decreasing. Hence as we move left from \hat{x} , even though the sign of the polarization effect is flipping to positive, the conglomeration effect is increasing in strength. Similarly

¹⁶ Keep in mind that we are still assuming $\mathbb{E}(g_j(\theta^j)) = 1$ for all j .

as we move to the right from $x(C)$, the conglomeration effect is just flipping to positive but the polarization effect is strictly negative and getting stronger.¹⁷ Moreover, the slopes of these effects (the second derivatives of U^1 and G_f) both depend on g'' . So any change to the functional form of g which increases the growth rate of the conglomeration effect in regions I and III would also increase the growth rate of the polarization effect. This suggests a stronger result on polar coalition formation. However, such a result is not forthcoming. For example, it is straight-forward to verify that condition can be violated for $g(z) = z^2$ and $\theta = 1$. Therefore, while the squared distance loss case represents an especially clean benchmark, it is not easy to generalize all of its implications.

4.4. Robustness Part 2: Equal Expected Private Wealth Assumption. How would relaxing the assumption on equal expected private wealth shares (3) affect whether condition (11) holds? Suppose, as one might expect, that centrists expect to receive a greater share of private wealth than extremists. Clearly, the polarization effect would be exaggerated and one would expect a violation of (11) in region I. But just how uneven should expected private wealth in continuation be? Whatever level of unevenness generated by a stationary equilibrium in an infinite horizon game is a natural benchmark. Could such a violation could be sustained in a stationary equilibrium in the infinite horizon game? Informal reasoning using the current framework suggests that such a violation would be difficult to sustain. If the polarization effect becomes too strong, it would lead to *polarized* (as opposed to polar) coalitions – coalitions which exclude centrists and are composed of both extreme right and extreme left members. However, such coalitions would give little or no cash to centrists thus destroying the basis for the exaggerated polarization effect in a stationary equilibrium. The next section takes a closer look at this question.

¹⁷ These observations are supported by equations (19) through (27) in the proof of Proposition 4.

POLAR COALITIONS IN THE INFINITE HORIZON GAME

In this section, I ask to what extent polar coalitions form the basis of stationary equilibria in an infinite horizon game. I derive a necessary and sufficient condition for polar coalition formation and present numerical analysis. There are three key differences between the model in this section and the previous section. First, the bargaining horizon is infinite. Second, n is even with $m = \frac{n}{2}$.¹⁸ Third, the configuration of ideal points is symmetric, so that if player $i \in N$ has ideal point x_i , there is some $j \in N$ with $x_j = 1 - x_i$. Fourth, cash shares are determined endogenously and so the assumption of equal expected cash shares (3) is dropped.

Define the coalitions L and R as the two polar coalitions with $L = \{0, 1, \dots, \frac{n}{2} - 1\}$ and $R = \{\frac{n}{2}, \frac{n}{2} + 1, \dots, n - 1\}$. To determine whether a pure strategy stationary equilibrium exists in which all players always choose to form polar coalitions, assume that all players always choose their own polar coalitions and then ask whether such a strategy is a best response. Proposition 5 offers a necessary and sufficient condition for polar equilibria.

Proposition 5. *If all players are expected to form the polar coalition to which they belong in the next round, it is a best response for any proposer $i \in L$ to form L in the current round if and only if inequality (14) holds for all $C \neq L$.*

$$(2 - \delta)[G(C) - G(L)] \geq \delta \left[\sum_{j \in C \cap L} g(|x_j - x(R)|) + \sum_{j \in C \cap R} g(|x_j - x(L)|) - \sum_{j \in L} g(|x_j - x(R)|) \right] \quad (14)$$

¹⁸ In the case of the infinite horizon version of this game, Jackson and Moselle (2002) show that, in general, pure strategy stationary equilibria do not exist when strict majorities are required. Since all potential proposers value low transfer demand, with strict majority coalitions and pure strategy profiles, it is impossible not to have at least one player - say the median - included in everyone's coalition. However, with a certain or near certain expectation of inclusion upon the failure of a proposal, the median player's transfer demand would have to be very high in equilibrium. This in turn gives proposers incentive not to include the median player thus breaking the equilibrium. The assumption that only $\frac{n}{2}$ players are needed for approval enables the existence of pure strategy stationary equilibria and allows one to focus on the polarization and conglomeration forces in the infinite horizon environment without introducing the complication of mixed strategies.

The case for a proposer in R is perfectly analogous. Just swap all occurrences of L and R (Proof: see Appendix).

The incentives of conglomeration and polarization are at work again. The left-hand side (LHS) of inequality (14) measures L 's closeness relative to C 's. If positive, it means that L is less spread out than C as measured by the total ideological loss felt by the coalition of implementing their own loss-minimizing policy. The right-hand side (RHS) of (14) measures C 's extremism relative to L 's. Extremism is good from the point of view of the proposer since it lowers transfer demands. A positive number on the RHS means that C is more extreme than L as measured by the total loss they feel when the opposite polar coalition is expected to win in the next round.¹⁹ Overall, (14) says that choosing L is a best response when L 's closeness (conglomeration) is at least as great as C 's extremism (polarization).

To see why polar equilibria often arise, as the numerical results below confirm, consider a coalition C that is different by just one member from L . Let $C = (L \setminus j) \cup q$ for some $j \in L$ and some $q \in R$. There are three possibilities. Either j is (i) equally, (ii) less or (iii) more extreme than player q . I will say that j is less extreme than q if $|x_j - \frac{1}{2}| < |x_q - \frac{1}{2}|$.

(i) Suppose that $x_q = 1 - x_j$ so that q and j are equally extreme. In this case, it is straightforward to show that the RHS of inequality (14) is exactly zero. Moreover, the LHS is clearly positive since C is more spread out than L as it is no longer contiguous. In short, L is more tightly packed than C and they are tied on extremism, making L the clear choice.

(ii) Suppose that q is more extreme than j . This makes the RHS of (14) positive and so the inequality can only hold if L can make up the difference in conglomeration. Note, however, that the more extreme q is compared to j , the larger is the conglomeration

¹⁹ Note that C is, by assumption a mix of members from L and R . C 's extremism is therefore, the first two summations terms inside the bracket, $\sum_{j \in C \cap R} g_j(R) + \sum_{j \in C \cap L} g_j(L)$, which are respectively the loss its left members expect to feel when R wins and the loss its right members feel when L wins.

effect. In other words, if we start from case (i) where $x_q = 1 - x_j$ and push x_q toward the right (holding x_j fixed), both sides of the inequality grow. Whether or not the RHS can grow quickly enough to break the inequality as q gets pushed further to the right depends on the parameters of the game, *but it has to overcome that initial deficit* observed in case (i). In this case, (14) is more likely to hold, for *smaller* δ .

(iii) Finally suppose that q is more moderate than j . This makes the RHS of the inequality unambiguously negative. However, the sign of the LHS is now ambiguous. Again, set x_q initially to $1 - x_j$, but this time push x_q to the left. Both sides of the inequality decrease as we make q more moderate. Whether or not C can break the inequality depends on whether it can overcome the initial deficit which, in general, depends on the parameters, but again it must overcome that deficit. In contrast to (ii), inequality (14) is more likely to hold for *larger* δ .

Summarizing these effects, when δ is small, it may be possible to break the equilibrium with a coalition which is tightly clustered but not necessarily pushed all the way to the left or the right. On the other hand, when δ is large it may be possible to break the equilibrium with a *polarized* coalition - one which takes its members from both extreme ends of the distribution, an arrangement which emphasizes extremism at the expense of closeness. So, consistent with the analysis of the previous section, urgency (low δ) may encourage coalitions of moderates while patience (high δ) may produce coalitions of extremists.²⁰ However, one must be careful. Just because a different type of coalition “breaks” the polar coalition equilibrium does not mean it necessarily represents a new pure strategy stationary equilibrium. It is entirely possible that the failure of inequality (14) to hold in either case (ii) or (iii) indicates the failure of any pure strategy stationary equilibrium to exist.

5.1. Numeric Trials. To test these predictions, I searched for polar coalition equilibria in games along the following dimensions.

- Legislature Size. All even integers $n \in [4, 18]$.

²⁰ Contrast this with the maxim of Banks and Duggan (2006) that patience implies moderation.

- Discount factor. All $\delta \in [0, 1]$ in increments of .02.
- Ideological loss function.
 - Squared Distance. $g(|x - x_i|) = (x - x_i)^2$
 - Linear. $g(|x - x_i|) = |x - x_i|$

For each parameter combination, I generated 10,000 games. Figure 4 summarizes the results with two diagrams, one for the case of the squared distance loss function and the other for the case of the linear loss function as indicated. In each diagram there are 51×8 cells representing each combination of n and δ outlined above. The color of each cell indicates the number of times out of 10,000 trials for the associated parameter combination that the polar coalition equilibrium described above was verified. Each of the 10,000 trials represented in each cell represents an independent random draw of symmetric ideal points (Figure 4).

There are two things to note here. First, in the vast majority of parameter combinations explored, the probability of a randomly drawn distribution of players' ideal points supporting polar coalition formation is extremely high. However, by no means, is it always equal to one. For example, the red regions toward in the left-hand diagram indicate that when the loss function is squared distance, virtually all trials with mid to upper values of δ had a stationary equilibrium in which all players form their own polar coalition.²¹ This is consistent with the theory from the previous section. There it was noted for the squared distance loss function, $\delta = 1$, and equal expected cash shares, polar coalitions were guaranteed (see Section 4.2). Here we have dispensed with the equal expected cash shares restriction which, as discussed in Section 4.4 should strengthen the polarization effect. Therefore, it is not surprising that, in order to maintain a virtual guarantee of polar coalition formation, we should have to dial down the discount factor; recall that lowering the discount factor strengthens the conglomeration effect and the two forces need to remain balanced to sustain polar coalition formation.

²¹ In fact, ALL the trials for $\delta \in [.68, .84]$ and squared distance loss had the polar coalition equilibrium as did all the trials for $\delta = 1$ and linear loss. The complete results as well as the java code which generated them are available upon request.

Second, the right-hand diagram in Figure 4 seems to indicate that making the loss function linear strengthens the conglomeration effect relative to the polarization effect since the regions of polar coalition equilibria is shifted up slightly toward higher discount factors. I do not offer a satisfying analytical explanation for this. As discussed in Section 4.3, there is no obvious reason to expect a change in the curvature of g to favor one incentive -conglomeration or polarization – over the other. I leave further investigation of this question to future research.

CONCLUSION

This paper takes the view that, fundamentally, legislatures bargain over ideological goods and the redistribution of wealth and that legislative outcomes are a product of limited compromise by members of winning coalitions along these dimensions. Thus, the question of *which* members belong to winning coalitions is important. The model used here, introduced to the literature by Jackson and Moselle (2002), has just enough detail to accomplish two goals. First, it allows the question of coalition composition by including players with heterogeneous preferences for an ideological good.²² Second, it allows for intra-coalition compromise by including money as another dimension of bargaining. As noted above, in the one-dimensional case without money analyzed by Banks and Duggan (2006), it is possible to ask about ideological composition of winning coalitions, but with the question becoming policy irrelevant as the discount factor approaches one. By contrast, the model used in this paper allows players to make trade-offs between ideological and distributional concerns that results in ideological policy choices toward the center of the winning coalition's ideal point distribution. Moderate coalitions choose moderate policy; extreme coalitions choose extreme policy.



So which kinds of coalition should we expect? This paper is an attempt to understand the forces shaping the ideological composition of winning coalitions and their subsequent policy choices. It is a difficult question to answer generally. I gained traction in this paper by focusing on polar coalitions and asking which game parameters supported them in equilibrium. The focus on polar coalitions has particular merits. Polar

²² The inability to make any meaningful queries about the composition of coalitions was acknowledged by Baron and Ferejohn themselves where they say, "The equilibrium strategies and distributions are thus unique up to the specification of who receives the benefits" (1989, p. 1187).

coalitions represent the most extreme ideal point configurations. As such, they choose the most extreme and least efficient policies.²³

My analysis isolates a pair of competing incentives in the coalition choice problem – conglomeration and polarization. Conglomeration is the incentive to have coalitions with members who are as close together ideologically as possible. It arises because members who are more distance from the central tendency of other members in a coalition demand more money to support the coalition’s policy choice. Polarization, on the other hand, is the incentive to include extremists. It arises because extremists have lower demand for shares of the private wealth due to their lower expected payoffs in continuation, *ceteris paribus*. Why? If there is uncertainty about the ideological policy in the event that bargaining continues to the next round, it is better to be a centrist.²⁴

The main conclusion of my analysis is that when these two incentives are balanced, conditions are ripe for polar coalition formation. As a first approximation, this can be understood intuitively. In the absence of conglomeration, polarization by itself would lead to coalitions of extremists from both ends of the ideological spectrum. On the other hand, conglomeration without polarization would lead to the most tightly packed configurations of ideal points regardless of their overall position in the ideological spectrum. Polar coalitions represent a compromise between these two forces. They are at once contiguous²⁵ and extreme.

The analysis culminating in Proposition 4 and illustrated in Figure 3 makes this more precise. When one imagines the proposer with the ability to move coalition members’ ideal points, one sees that conglomeration pulls coalition members in toward the *coalition* center, , while polarization pushes them out away from the *population* center, . Between these two points, the two incentives agree that members should

²³ For example, in the case of squared distance loss functions, the policy that minimizes the sum of ideological loss across all players would be the mean of all players’ ideal points. However, coalitions choose the policy equal to the mean taken only over their members’ ideal points and polar coalitions have the most extreme means.

²⁴ This fact is especially obvious when the loss function is convex (as assumed in this model). However, readers can easily check that centrists do better than extremists on average even in a world where everyone’s loss function is concave.

²⁵ A contiguous ideal point configuration is one made up of adjacent ideal points from the population of ideal points. The most closely packed coalition would be contiguous, but not all contiguous coalitions would be the most closely packed.

move in the direction that the coalition already leans – in other words, toward the polar configuration. As we move outside the region of agreement in either direction, it is always the incentive that would push members toward the polar configuration that has a head start and that one generally expects to dominate. However, there is no guarantee under all game parameters. If something in the model favors one of the two incentives over the other, then the formation of polar coalitions will likely depend on the distribution of ideal points. I also discussed how the sufficient and necessary condition from Proposition 5 extends this logic to the infinite horizon game.

In both the finite and infinite environments, I showed how increasing the discount factor strengthens polarization relative to conglomeration. This naturally suggests the following hypothesis: If polar coalition formation in equilibrium requires balance between the two forces and the discount factor determines their relative weight, then there exists a discount factor or range of discount factors which more or less guarantees polar coalition formation regardless of the distribution of ideal points. Indeed, for the generalized ultimatum game with squared distance loss, I showed that setting $\delta = 1$ does exactly that. Moreover, in the infinite horizon game, I confirm ranges of δ that virtually guarantee polar coalition formation for every game size up to 18 players and two distinct loss functions.

The strategic environment modeled in this paper is very spare compared to real world institutions such as the U.S. Congress, European parliaments or state legislatures. The model contains a single ideological dimension, proposals are voted up or down in single-shot game (though allowing for possibly many rounds of bargaining) and there is no explicit account of political parties. By contrast, the strategic environment of real-world legislatures may feature several ideological dimensions, multi-layer decision making in committees, reconciliation of bills between multiple “houses”, an agenda setting process, repeated interactions and reputation effects, party affiliation, and so on. These are all important institutional considerations that may well affect the ideological structure of winning coalitions. Perhaps the most interesting of these issues to explore in future research is the explicit role of party affiliation. However, since theoretical analysis must always trade-off between realism and tractability, I set those questions aside.

Though stripped down relative to the real world, the model is still relevant for two reasons. First, the model identifies and explains plausible sources of the conglomeration and polarization incentives even if these incentives may have to compete with others in real-world legislatures. Second, one need not believe the model's explanation for the origin of the conglomeration and polarization incentives in order to accept the logic that the combination of these forces leads to polar coalition formation. One need only believe that these are real incentives for *some* reason. With regard to conglomeration, one could easily imagine, for example, that like-minded members of legislatures develop closer friendships and therefore find it easier to vote together. Similarly with polarization, extremists could occupy weak bargaining positions relative to moderates for a variety of reasons. For example, in a richer environment with repeated interactions and reputation effects, it seems reasonable that moderates could more credibly threaten to vote with the other side in future bargaining sessions. Conversely, when one expects the forces to be lopsided, this model offers a way to think about why to expect a lower probability of polar coalitions. For example, perhaps the only issue is ideological and the opportunity for redistribution of wealth among the players as part of the bargaining process is constrained (e.g., judicial panels and faculty committees) thereby removing the mechanism that sustains polarization.

APPENDIX A. PROOFS

Proof of Proposition 1. I analyze a simplified voting game which may be thought of as a truncated subgame in the larger bargaining game. This simple voting game has n players labeled $(0, \dots, n - 1)$ who vote YES or NO in sequence. If a critical number, $m > 0$, vote YES then each player i gets payoff $u_i(\text{PASS})$. Otherwise each player i gets $u_i(\text{FAIL})$. Let A be the subset of player who strictly prefer PASS to FAIL. Let D be the subset of players who strictly prefer FAIL to PASS.

Players vote sequentially with each player observing the votes of those who vote earlier. A pure strategy for player i is a plan of action - to vote YES or NO - at each of its $2i$ decision nodes. A mixed strategy assigns to each node a probability of voting YES. A

subgame perfect equilibrium requires that all players in A have a strategy which prescribes voting YES with probability one whenever doing so improves the chances of passage. Similarly, (though a bit counters intuitively) it also requires that members of A vote YES with probability zero whenever voting YES diminishes the chances of passage. Opposite statements apply to D .

We want to show that any time there is a majority of players in A , then the motion will pass with probability one in any subgame perfect equilibrium. Assume $\#A \geq m$. Create a new index for the players belonging to A with labels $(k_0, \dots, k_{\#A-1})$ and assign these labels in reverse order of these players' rank in the voting sequence. Hence k_0 is the last player to vote among those who are strictly in favor of the proposal. Also player $k_{\#A-1}$ is the player who is strictly in favor of the proposal and who would be pivotal, if among all other players after him who are strictly in favor vote YES and all those before him voted NO. Note that player k_0 has a non-empty set of decision nodes from which a YES vote guarantees passage of the proposal no matter how voters following him in the voting sequence vote. Call these k_0 's guarantee nodes. For example, the decision node or nodes reached when at least $m-1$ of the players who are strictly in favor of the proposal have all voted YES must be among k_0 's guarantee nodes. By the definition of a subgame perfect equilibrium, if play reaches any of these nodes, either player k_0 must vote YES or the chances of the proposal passing after a NO vote by player k_0 must be equal to one as well. Therefore, the proposal must pass with probability one if play reaches any of k_0 's guarantee nodes in any SPE.

Now consider the nodes belonging to the second to last voter strictly in favor of the motion, player k_1 . A non-empty subset of decision nodes belongs to this player from which a YES vote invariably leads to at least one of k_0 's guarantee nodes. Call these k_1 's guarantee nodes. In general, define k_i 's guarantee nodes to be those decision nodes from which k_i has the power to direct play invariably to one of k_{i-1} 's guarantee nodes. It is clear that for all i , all the decision nodes at player k_i 's turn reached after $(m-1) - i$ players preceding him in the voting sequence have voted YES are guarantee

nodes for player k_i . It follows that all of player k_{m-1} 's decision nodes are guarantee nodes. In particular player k_{m-1} can guarantee that play will reach player k_{m-2} 's guarantee nodes with a YES vote from any of his nodes. Therefore whether player k_{m-1} votes YES or not, the proposal must pass in any SPE with probability one following his vote. A symmetric argument demonstrates that the proposal will fail whenever a majority strictly prefers it to fail. QED.

Proof of Proposition 2. The proof addresses the proposition point-by-point.

(i) To show that $i \in C(\theta^0)$ we need to show that i can construct a passable proposal $\tilde{\theta}$ with $u_i(\tilde{\theta}) > \delta E u_i(\theta^1)$. Note that $\tilde{\theta}$ need not be θ^0 . To construct an example of one such $\tilde{\theta}$, randomly choose $\frac{n-1}{2}$ players not i . Call this subset along with i (optimistically) $C(\tilde{\theta})$. Now set $x(\tilde{\theta}) = E(x(\theta^1))$ and $y_j(\tilde{\theta}) = \frac{n}{n-1}$ for all $j \in C(\tilde{\theta})$ and $y_k(\tilde{\theta}) = 0$ for all $k \notin C(\tilde{\theta})$. To show that $\tilde{\theta}$ would pass in any SPE and that all $j \in C(\tilde{\theta})$ (including i) would be in favor, we need only show that $u_j(\tilde{\theta}) > \delta E u_j(\theta^1)$ for all $j \in C(\tilde{\theta})$. Note that

$$E u_j(\theta^1) = 1 - \int_0^1 g(|x_j - x(\theta^1)|) dF(x(\theta^1))$$

while for all $j \in C(\tilde{\theta})$

$$u_j(\tilde{\theta}) = \frac{n}{n-1} - g(|x_j - E x(\theta^1)|)$$

But $\frac{n}{n-1} > 1$ and $g'' > 0 \Rightarrow g(|x_j - E x(\theta^1)|) < \int_0^1 g(|x_j - x(\theta^1)|) dF(x(\theta^1))$. Therefore,

$u_j(\tilde{\theta}) > E u_j(\theta^1)$ for all $j \in C(\tilde{\theta})$ even if $\delta = 1$. Therefore by Proposition 1, the proposal would pass in any subgame perfect equilibrium.

(ii) Equation (4) specifies an amount of cash which makes any member of $C(\theta^0)$ indifferent between θ^0 and continuation. The assumptions that $E y_j(\theta^1) = 1$ for all $j \in N$, $g(1) = 1$ and that $F(x(\theta^1))$ puts at least some weight on two policies imply that $E u_j(\theta^1) > 0$ for all $j \in N$ and therefore $y_k(\theta^0) > 0$ for all $k \in C(\theta^0)$. The claim that $\#C(\theta^0) = m$ follows directly from the strictly positive cash shares of all coalition members, since to have more players in the coalition would unnecessarily lower the proposer's payoff.

(iii) The uniqueness of $x(\theta^0)$ and therefore of θ^0 for a given C in favor follows from the assumption that $g'' > 0$ and the fact that the sum of convex functions is also convex.

(iv) Proposer, i 's payoff is computed by observing that i is the residual claimant on the pot of cash and so $y_i(\theta^0) = n - \sum_{j \in C(\theta^0)} y_j(\theta^0)$. Applying this to (1) and substituting in the expression for coalition member cash shares from (4) yields the result. QED.

Proof of Proposition 3. In order to more explicitly present $x(C_{z_a})$ as a function of the replacement member's ideal point, x_{z_a} , let $x(x_{z_a})$ denote $x(C_{z_a})$ from above. Because, by definition, $x(x_a)$ solves the minimization problem from (5), it must satisfy the associated first-order condition. Hence when $x_{z_a} = x_a$, we have

$$\sum_{j|x_j < x(x_a)} g'(x(x_a) - x_j) = \sum_{j|x_j > x(x_a)} g'(x_j - x(x_a)) \quad (15)$$

Suppose that $x_a < x(x_a)$. (The logic for the other case is symmetric.) Set $x_{z_a} = x_a + \Delta$ with $\Delta > 0$ small enough so that z_a 's ideal point has the same rank as a 's and still lies below $x(x_a)$. This proof will, in essence, show that $x(x_a) < x(x_a + \Delta) = x(x_{z_a}) < x(x_a) + \Delta$. Just as (15) must hold with respect to coalition C , equation (16) must hold for C_{z_a} :

$$g'(x(x_a + \Delta) - (x_a + \Delta)) + \sum_{j \in C|x_j < x(x_a)} g'(x(x_a + \Delta) - x_j) = \sum_{j \in C|x_j > x(x_a)} g'(x_j - x(x_a + \Delta)) \quad (16)$$

Note that x_{z_a} is written explicitly as $x_a + \Delta$ and I've separated out the a th term, which belongs on the LHS by our supposition that $x_a < x(C)$. Subtracting (15) from (16) we get

$$\begin{aligned}
& g'(x(x_\alpha + \Delta) - (x_\alpha + \Delta)) - g'(x(x_\alpha) - x_\alpha) \\
& + \sum_{j \in C_j | j \neq \alpha, x_j < x(x_\alpha)} [g'(x(x_\alpha + \Delta) - x_j) - g'(x(x_\alpha) - x_j)] \\
& = \sum_{j \in C_j | x_j > x(x_\alpha)} [g'(x_j - x(x_\alpha + \Delta)) - g'(x_j - x(x_\alpha))]
\end{aligned}
\tag{17}$$

In the limit as $\Delta \rightarrow 0$ by the definition of a derivative, we can write

$$x(x_\alpha + \Delta) = x(x_\alpha) + x'(x_\alpha)\Delta.$$

Substituting this back into our difference equation (17) we get

$$\begin{aligned}
& g'(x(x_\alpha) + x'(x_\alpha)\Delta - (x_\alpha + \Delta)) - g'(x(x_\alpha) - x_\alpha) \\
& + \sum_{j \in C_j | j \neq \alpha, x_j < x(x_\alpha)} [g'(x(x_\alpha) + x'(x_\alpha)\Delta - x_j) - g'(x(x_\alpha) - x_j)] \\
& = \sum_{j \in C_j | x_j > x(x_\alpha)} [g'(x_j - x(x_\alpha) - x'(x_\alpha)\Delta) - g'(x_j - x(x_\alpha))]
\end{aligned}
\tag{18}$$

Using the definition of a derivative again, we can write down the following three identities in the limit as $\Delta \rightarrow 0$.

$$g'(x(x_\alpha) + x'(x_\alpha)\Delta - (x_\alpha + \Delta)) - g'(x(x_\alpha) - x_\alpha) = g''(x(x_\alpha) - x_\alpha)\Delta(x'(x_\alpha) - 1)$$

$$g'(x(x_\alpha) + x'(x_\alpha)\Delta - x_j) - g'(x(x_\alpha) - x_j) = g''(x(x_\alpha) - x_j)x'(x_\alpha)\Delta$$

$$g'(x_j - x(x_\alpha) - x'(x_\alpha)\Delta) - g'(x_j - x(x_\alpha)) = -g''(x_j - x(x_\alpha))x'(x_\alpha)\Delta$$

Using these we can re-write (18) as

$$\begin{aligned}
& g''(x(x_\alpha) - x_\alpha)\Delta(x'(x_\alpha) - 1) \\
& + \sum_{j \in C_j | j \neq \alpha, x_j < x(x_\alpha)} [g''(x(x_\alpha) - x_j)x'(x_\alpha)\Delta] = \sum_{j \in C_j | x_j > x(x_\alpha)} [-g''(x_j - x(x_\alpha))x'(x_\alpha)\Delta]
\end{aligned}$$

Canceling the Δ s and absorbing what we can of the α th term back into the sum, we get

$$-g''(x(x_\alpha) - x_\alpha) + \sum_{j \in C} g''(|x(x_\alpha) - x_j|)x'(x_\alpha) = 0$$

$$\Rightarrow x'(x_a) = \frac{g''(x(x_a) - x_a)}{\sum_{j \in C} g''(|x(x_a) - x_j|)}$$

which obviously lies between 0 and 1.

Proof of Proposition 4. The proof proceeds as follows.

- (1) Establish several key characteristics of G , G_f and U^1 justifying Figures 1 and 2.
- (2) Prove the equivalency of moving a coalition member's ideal point and swapping the member.
- (3) Prove a cumbersome version of the sufficient condition involving G instead of G_f .
- (4) Simplify the sufficient condition by proving that the bias introduced by replacing G with G_f always works in the right direction.

Part (1). Player a 's individual conglomeration effect is given by

$$\begin{aligned} \frac{dG(C_{a_n})}{dx_{a_n}} &= \frac{d}{dx_{a_n}} \sum_{j \in C_{a_n}} g(|x_{a_n} - x(C_{a_n})|) \\ &= \frac{\partial g(|x_{a_n} - x(C_{a_n})|)}{\partial x_{a_n}} + \sum_{j \in C} \frac{\partial g(|x_{a_n} - x(C_{a_n})|)}{\partial x(C_{a_n})} \frac{\partial x(C_{a_n})}{\partial x_{a_n}} \\ &= \frac{\partial g(|x_{a_n} - x(C_{a_n})|)}{\partial x_{a_n}} \\ &= g'(|x_{a_n} - x(C_{a_n})|) \text{sgn}(x_{a_n} - x(C_{a_n})) \end{aligned}$$

(19)

The equality between the second and third line holds because $x(C)$, by definition, solves the coalition ideological loss minimization problem, meaning that the terms in the summation add to zero. Since $x(C)$ is held constant in the definition of G_f , coalition C 's field conglomeration effect is

$$\frac{dG_f(C_{a_n})}{dx_{a_n}} = g'(|x_{a_n} - x(C)|) \text{sgn}(x_{a_n} - x(C))$$

(20)

The slope (2nd derivative) of a 's individual conglomeration effect is given by

$$\begin{aligned} \frac{d^2 G(C_{\alpha})}{dx_{\alpha}^2} &= \frac{\partial^2 g(|x_{\alpha} - x(C_{\alpha})|)}{\partial x_{\alpha}^2} + \frac{\partial^2 (|x_{\alpha} - x(C_{\alpha})|)}{\partial x_{\alpha} \partial x(C)} \frac{\partial x(C)}{\partial x_{\alpha}} \\ &= g''(|x_{\alpha} - x(C)|) \left(1 + \operatorname{sgn}(x_{\alpha} - x(C)) \operatorname{sgn}(x(C) - x_{\alpha}) \frac{\partial x(C)}{\partial x_{\alpha}} \right) \\ &= g''(|x_{\alpha} - x(C)|) \left(1 - \frac{\partial x(C)}{\partial x_{\alpha}} \right) \end{aligned} \quad (21)$$

The slope of C 's field conglomeration effect:

$$\begin{aligned} \frac{d^2 G_f(C_{\alpha})}{dx_{\alpha}^2} &= g''(|x_{\alpha} - x(C)|) [\operatorname{sgn}(x_{\alpha} - x(C))]^2 \\ &= g''(|x_{\alpha} - x(C)|) \end{aligned} \quad (10)$$

The signs of and relationship between field and individual conglomeration effect slopes:

$$0 < \frac{d^2 G(C_{\alpha})}{dx_{\alpha}^2} < \frac{d^2 G_f(C_{\alpha})}{dx_{\alpha}^2} \quad (23)$$

Inequality (23) follows from (22) and (21) and the fact that

$$0 < \frac{dx(C)}{dx_{\alpha}} < 1 \quad (24)$$

which is given by Proposition 3. Equations (19) and (20) and the fact that $x(C_{\alpha})|_{x_{\alpha}=x_{\alpha}} = x(C)$ immediately imply that C 's field and a 's individual conglomeration effects are equal when evaluated at $x_{\alpha} = x_{\alpha}$.

$$\left. \frac{dG(\mathcal{C}_{\mathcal{B}_a})}{dx_{\mathcal{B}_a}} \right|_{x_{\mathcal{B}_a}=x_a} - \left. \frac{dG(\mathcal{C}_{\mathcal{B}_a})}{dx_{\mathcal{B}_a}} \right|_{x_{\mathcal{B}_a}=x_a} \quad (25)$$

The polarization effect (the first derivative of U^1) is independent of α and \mathcal{C} . It reduces to the following.

$$\frac{dU^1(\mathcal{C}_{\mathcal{B}_a})}{dx_{\mathcal{B}_a}} = - \int_0^{x_{\mathcal{B}_a}} g'(x_{\mathcal{B}_a} - x^1) dF(x^1) + \int_{x_{\mathcal{B}_a}}^1 g'(x^1 - x_{\mathcal{B}_a}) dF(x^1) \quad (26)$$

Its second derivative is strictly negative and is given by the following.

$$\frac{d^2U^1(\mathcal{C}_{\mathcal{B}_a})}{dx_{\mathcal{B}_a}^2} = - \int_0^{x_{\mathcal{B}_a}} g''(x_{\mathcal{B}_a} - x^1) dF(x^1) - \int_{x_{\mathcal{B}_a}}^1 g''(x^1 - x_{\mathcal{B}_a}) dF(x^1) - 2g'(0)F'(x_{\mathcal{B}_a}) < 0 \quad (27)$$

Both (26) and (27) follow directly from the definition of $U^1(\mathcal{C})$ and Leibniz's Rule.

Part (2). Consider two coalitions of size m , \mathcal{C}' and \mathcal{C}'' , which are different by just one member. \mathcal{C}' includes player a while \mathcal{C}'' has b , with $\#(\mathcal{C}' \cap \mathcal{C}'') = m - 1$, $b \in \mathcal{C}'$ and $a \in \mathcal{C}''$. By the definition of $\mathcal{C}_{\mathcal{B}_a}^i$ it is obvious that

$$\begin{aligned} u_t(\mathcal{C}_{\mathcal{B}_a}^i) \Big|_{x_{\mathcal{B}_a}=x_a} &= u_t(\mathcal{C}') \\ u_t(\mathcal{C}_{\mathcal{B}_a}^i) \Big|_{x_{\mathcal{B}_a}=x_b} &= u_t(\mathcal{C}'') \end{aligned}$$

Therefore (28) is the difference in payoff to i of choosing coalition \mathcal{C}'' over \mathcal{C}' .

$$\begin{aligned} u_t(\mathcal{C}'') - u_t(\mathcal{C}') &= u_t(\mathcal{C}_{\mathcal{B}_a}^i) \Big|_{x_{\mathcal{B}_a}=x_b} - u_t(\mathcal{C}_{\mathcal{B}_a}^i) \Big|_{x_{\mathcal{B}_a}=x_a} \\ &= [G(\mathcal{C}_{\mathcal{B}_a}^i) + \partial U^1(\mathcal{C}_{\mathcal{B}_a}^i)] \Big|_{x_{\mathcal{B}_a}=x_b} - [G(\mathcal{C}_{\mathcal{B}_a}^i) + \partial U^1(\mathcal{C}_{\mathcal{B}_a}^i)] \Big|_{x_{\mathcal{B}_a}=x_a} \quad (28) \\ &= \int_{x_{\mathcal{B}_a}=x_a}^{x_b} \left[\frac{d}{dx_{\mathcal{B}_a}} (G(\mathcal{C}_{\mathcal{B}_a}^i) + \partial U^1(\mathcal{C}_{\mathcal{B}_a}^i)) \right] dx_{\mathcal{B}_a} \end{aligned}$$

The transition from the first line to the second follows from (7) while the last line is a straight-forward application of the fundamental theorem of calculus.

Part (3). Obviously if the sign of the integrand in equation (28) is positive (negative) over the entire interval $[x_a, x_b]$ then C'' is an unambiguously better (worse) choice for i than C' . Hence we can clearly make the following (slightly cumbersome) claims. If, for all choices of $a \in C$, whenever $\hat{x} - x(C) < 0$, we have (29) and whenever $\hat{x} - x(C) > 0$, we have (30) then polar coalition formation is required in any subgame perfect equilibrium.

$$\frac{d}{dx_{za}} (G(C_{za}) + \delta U^1(C_{za})) < 0 \quad \forall x_{za} \in [x_a, 1] \quad (29)$$

$$\frac{d}{dx_{za}} (G(C_{za}) + \delta U^1(C_{za})) > 0 \quad \forall x_{za} \in [0, x_a] \quad (30)$$

To see why, suppose that $\hat{x} - x(C) < 0$, so that C is a right-leaning coalition. The claim is that an even more right-leaning coalition would reduce the sum of G and U^1 (and therefore increase u_i) which would obviously be true if inequality (29) held and there exists some player $b \in C$ with $b \neq i$ and $x_b > x_a$. If there does not exist such a player b , then $C = R_i$. (That is, C already is i 's right polar coalition.) The argument is symmetric for the case of a left-leaning coalition. Therefore, if both conditions held true, L_i would be the best choice among all left-leaning coalitions and R_i would be the best choice among all right-leaning coalitions.

Part (4). To prove the simpler version of the sufficient condition stated in the proposition, we must show that replacing G with G_f will not flip the signs of inequalities (29) and (30) when, respectively, $\hat{x} - x(C) < 0$ and $\hat{x} - x(C) > 0$. To show this, it suffices to prove the following two statements.

$$\hat{x} - x(C) < 0 \Rightarrow \frac{d}{dx_{za}} G(C_{za}) \leq \frac{d}{dx_{za}} G_f(C_{za}) \quad \forall x_{za} \in [x_a, 1]$$

$$x - x(C) > 0 \Rightarrow \frac{d}{dx_{a_n}} G(C_{a_n}) \geq \frac{d}{dx_{a_n}} G_f(C_{a_n}) \forall x_{a_n} \in [x_{a_n}, 1]$$

These two statements follow from the observations made above in (25) and (23). (Note: Figure 4.1 illustrates these points) QED.

Proof of Proposition 5. To save space in this proof I use the following shorthand. Let $g_j(C) \equiv g(|x_j - x(C)|)$. The sufficient and necessary condition (14) at the heart of the Proposition can therefore be re-written as follows.

$$(2 - \delta)[G(C) - G(L)] \geq \delta \left[\sum_{j \in C \cap L} g_j(R) + \sum_{j \in C \cap R} g_j(L) - \sum_{j \in L} g_j(R) \right] \quad (31)$$

The proof proceeds as follows.

- (1) Derive the solution for each player's expected utility in equilibrium.
- (2) Set up the coalition choice problem.

Part (1). *Solution to System of Expected Utilities Equations.* In a stationary equilibrium where everyone always chooses the polar coalition to which that player belongs, the expected utility of player $i \in L$ is given by the following.

$$\begin{aligned} E(u_i) = & \frac{n-2}{2n} \delta E(u_i) \\ & + \frac{1}{n} \left[n - \sum_{j \in L \setminus i} \delta E(u_j) - \sum_{j \in L} g_j(L) \right] \\ & - \frac{1}{2} g_i(R) \end{aligned} \quad (32)$$

On the first line, $\frac{n-2}{2n}$ is the chance that someone in L other than i will be selected as the proposer and will have to give i his continuation value $\delta E(u_i)$. On the second line, $\frac{1}{n}$ is the probability that i is selected as the proposer and the terms in the square brackets represent his payoff in that case. On the third line, $\frac{1}{2}$ is the chance that someone from

the other polar coalition, R , is selected to be the proposer in which case i loses $g_i(R)$.

Collecting terms we have

$$\left(\frac{2n + 2\delta - \delta n}{2n}\right)E(u_i) + \frac{\delta}{n} \sum_{j \in L \setminus i} E(u_j) = 1 - \frac{1}{n} \sum_{j \in L} g_j(L) - \frac{1}{2} g_i(R)$$

Multiplying thru by $\frac{2n}{2n + 2\delta - \delta n}$ we get

$$E(u_i) + \frac{2\delta}{2n + 2\delta - \delta n} \sum_{j \in L \setminus i} E(u_j) = \frac{1}{2n + 2\delta - \delta n} \left(2n - 2 \sum_{j \in L} g_j(L) - n g_i(R) \right)$$

This forms a linear system of $m = \frac{n}{2}$ equations which, in matrix notation, is

$$\begin{pmatrix} 1 & \alpha & \alpha & \dots & \alpha \\ \alpha & 1 & \alpha & \dots & \alpha \\ \dots & \dots & \dots & \dots & \dots \\ \alpha & \dots & \alpha & \alpha & 1 \end{pmatrix} \begin{pmatrix} E(u_0) \\ E(u_1) \\ \dots \\ E(u_{\frac{n}{2}-1}) \end{pmatrix} = \begin{pmatrix} b_0 \\ b_1 \\ \dots \\ b_{\frac{n}{2}-1} \end{pmatrix}$$

where $\alpha = \frac{2\delta}{2n + 2\delta - \delta n}$ and for each $i \in L$

$$b_i = \frac{1}{2n + 2\delta - \delta n} \left(2n - 2 \sum_{j \in L} g_j(L) - n g_i(R) \right).$$

It is easily verified that a system of linear equations with this form has the following solution

$$E(u_i) = \frac{((m-2)\alpha + 1)b_i}{1 + (m-2)\alpha - (m-1)\alpha^2} + \frac{\alpha \sum_{j \in L \setminus i} b_j}{(m-1)\alpha^2 - (m-2)\alpha - 1}$$

Putting α , m and b_i back in terms of δ , n and $g_i(L)$ and $g_i(R)$, we get

$$E(u_i) = \frac{1}{(\delta - 2)n^2} \left[(n - \delta) \left(2 \sum_{j \in L} g_j(L) + n g_i(R) - 2n \right) - \sum_{k \in L \setminus i} \delta \left(2 \sum_{j \in L} g_j(L) + n g_k(R) - 2n \right) \right]$$

which reduces to

$$E(u_i) = 1 - \frac{1}{n} \sum_{j \in L} g_j(L) - \frac{g_i(R)}{2 - \delta} + \frac{\delta}{n(2 - \delta)} \sum_{j \in L} g_j(R) \quad (33)$$

Swapping R and L in this expression yields the solution for the expected utilities of members of R .

Part (2). For any $i \in L$, L is clearly a better choice than C whenever we have

$$n - \delta \sum_{j \in L} E(u_j) - \sum_{j \in L} g_j(L) > n - \delta \sum_{j \in C} E(u_j) - \sum_{j \in C} g_j(C)$$

The LHS is proposer i 's payoff from choosing L in equilibrium (assuming it exists). The RHS is i 's payoff from choosing some arbitrary coalition C . Note that $i \in L$ by assumption. However, we can also assume that $i \in C$ since proposers never want to choose anything but minimum majority coalitions for the same reason discussed in the proof of Proposition 2. Subtract n , multiply by -1 , and add $\delta E(u_i)$ to both sides of this inequality to get

$$\sum_{j \in L} (\delta E(u_j) + g_j(L)) \leq \sum_{j \in C} (\delta E(u_j) + g_j(C)) \quad (34)$$

The quantities on the left and right hand sides of (34) are the total bribe demands for coalitions L and C respectively. Substitute the expression for expected utility from (33) into (34) for members of L and the analogous expression for members of R to get

$$\begin{aligned} & \sum_{j \in L} \left(\delta \left[1 - \frac{1}{n} \sum_{k \in L} g_k(L) - \frac{g_j(R)}{2 - \delta} + \frac{\delta}{n(2 - \delta)} \sum_{k \in L} g_k(R) \right] + g_j(L) \right) < \\ & \sum_{j \in C \cap L} \left(\delta \left[1 - \frac{1}{n} \sum_{k \in L} g_k(L) - \frac{g_j(R)}{2 - \delta} + \frac{\delta}{n(2 - \delta)} \sum_{k \in L} g_k(R) \right] + g_j(C) \right) + \\ & \sum_{j \in C \cap R} \left(\delta \left[1 - \frac{1}{n} \sum_{k \in R} g_k(R) - \frac{g_j(L)}{2 - \delta} + \frac{\delta}{n(2 - \delta)} \sum_{k \in R} g_k(L) \right] + g_j(C) \right) \end{aligned}$$

Note that I have decomposed the summation on the RHS into members of $C \cap L$ and $C \cap R$. (Since n is even L and R are exhaustive and non-overlapping.) This is necessary because, as noted in Part (1), the expressions for the expected utility of members of L and members of R are different (though symmetrically analogous). Now I make three cancellations. First there are the same number of δ terms on each side of the

inequality. Second, by the assumption that the ideal point configuration is symmetric, $\sum_{k \in L} g_k(R) = \sum_{k \in R} g_k(L)$, and there are the same number of these terms on each side of the inequality. Third, again by the symmetry of the ideal points, $\sum_{k \in L} g_k(L) = \sum_{k \in R} g_k(R)$, and there are the same number of these terms on each side. Hence, the inequality reduces to

$$\sum_{j \in L} \left(g_j(L) - \delta \left[\frac{g_j(R)}{2 - \delta} \right] \right) < \sum_{j \in C \cap L} \left(g_j(C) - \delta \left[\frac{g_j(R)}{2 - \delta} \right] \right) + \sum_{j \in C \cap R} \left(g_j(C) - \delta \left[\frac{g_j(L)}{2 - \delta} \right] \right)$$

The transition to (31) follows from the fact that

$$G(L) = \sum_{j \in L} g_j(L) \quad \text{and} \quad G(C) = \sum_{j \in C} g_j(C) = \sum_{j \in C \cap L} g_j(C) + \sum_{j \in C \cap R} g_j(C). \quad \text{QED.}$$

ACKNOWLEDGMENTS

I would to thank Larry Samuelson, William Sandholm, Peter Norman and Jim Andreoni for many helpful remarks as I worked out the theory on early versions of this paper. Many others including Kara Leibel, Peter Grajzl, Jo Hertel and Laura Schecter generously read drafts and provided comments. Errors, over sights and omissions are all mine. "To date, game theory has been unable, in general, to explain or predict *which* coalitions will emerge when a game is played. This surprising gap needs to be closed ..." - Beth Allen (2000), my italics.

REFERENCES

- Alesina, A., & Rosenthal, H. (2000). Polarized platforms and moderate policies with checks and balances. *Journal of Public Economics*, 75(1), 1-20.
- Allen, B. (2000). The future of microeconomic theory. *The Journal of Economic Perspectives*, 14(1), 143-150.
- Austen-Smith, D., & Banks, J. (1988). Elections, coalitions, and legislative outcomes. *The American Political Science Review*, 82(2), 405-422.
- Banks, J. S., & Duggan, J. (2000). A bargaining model of collective choice. *The American Political Science Review*, 94(1), 73-88.
- Banks, J.S., & Duggan, J. (2006). A general bargaining model of legislative policy-making. *Quarterly Journal of Political Science*, 1(1), 49-85.
- Baron, D. (1991). A spatial bargaining theory of government formation in parliament systems. *American Political Science Review*, 85, 137-164.
- Baron, D. (1996). A dynamic theory of collective goods programs. *American Political Science Review*, 90(2), 316-330.
- Baron, D. P., & Diermeier, D. (2001). Elections, governments and parliaments in proportional representation systems. *Quarterly Journal of Economics*, 116(3), 933-967.
- Baron, D. P., Diermeier, D., & Fong, P. (2007). Policy dynamics and inefficiency in a parliamentary democracy with proportional representation. SSRN eLibrary.
- Baron, D. P., & Ferejohn, J. A. (1989). Bargaining in legislatures. *The American Political Science Review*, 83(4), 1181-1206.
- Battaglini, M., & Coate, S. (2007). Inefficiency in legislative policymaking: A dynamic analysis. *The American Economic Review*, 97(1), 118-149.

- Battaglini, M., & Coate, S. (2008). A dynamic theory of public spending, taxation, and debt. *American Economic Review*, 98(1), 201-236.
- Calvert, R. L., & Dietz, N. (1996). Legislative coalitions in a bargaining model with externalities. *University of Rochester Working Paper*.
- Harrington, J. E. (1989). The advantageous nature of risk aversion in a three-player bargaining game where acceptance of a proposal requires a simple majority. *Economics Letters*, 30(3), 195-200.
- Ingberman, D., & Villani, J. (1993). An institutional theory of divided government and party polarization. *American Journal of Political Science*, 37(2), 429-471.
- Jackson, M. O., & Moselle, B. (2002). Coalition and party formation in a legislative voting game. *Journal of Economic Theory*, 103(1), 49-87.
- Krehbiel, K., (2000). Party Discipline and Measures of Partisanship. *American Journal of Political Science*, 44(2), 212-227.
- Lizzeri, A., & Persico, N. (2001). The provision of public goods under alternative electoral incentives. *The American Economic Review*, 91(1), 225-239.
- McCarty, N., Poole, K., & Rosenthal H. (2001). The Hunt for Party Discipline in Congress. *American Political Science Review*, 95(3), 673-687.
- Norman, P. (2002). Legislative bargaining and coalition formation. *Journal of Economic Theory*, 102(2), 322-353.
- Palfrey, T. R. (1984). Spatial equilibrium with entry. *The Review of Economic Studies*, 51(1), 139-156.
- Poole, K.T. & Rosenthal H. (1991). Patterns of Congressional Voting. *American Journal of Political Science*, 35(1), 228-278.
- Romer, T., & Rosenthal, H. (1979). Bureaucrats versus voters: On the political economy of resource allocation by direct democracy. *The Quarterly Journal of Economics*, 93(4), 563-587.
- Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. *Econometrica*, 50(1), 97-109.
- Selten, R. (1975). Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory*, 4(1), 25-55.
- Volden, C., & Wiseman, A. E. (2007). Bargaining in legislatures over particularistic and collective goods. *American Political Science Review*, 101(1), 79-92.

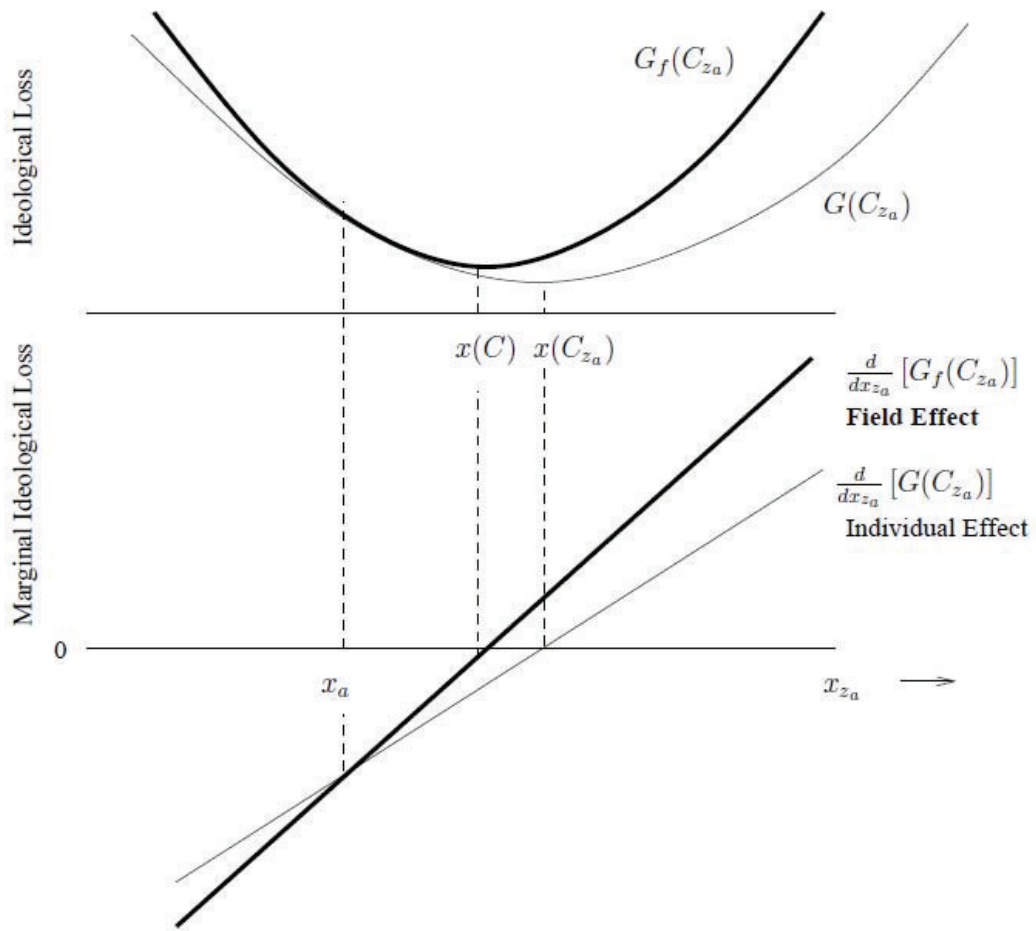


Figure 1. Field and individual conglomeration effects.

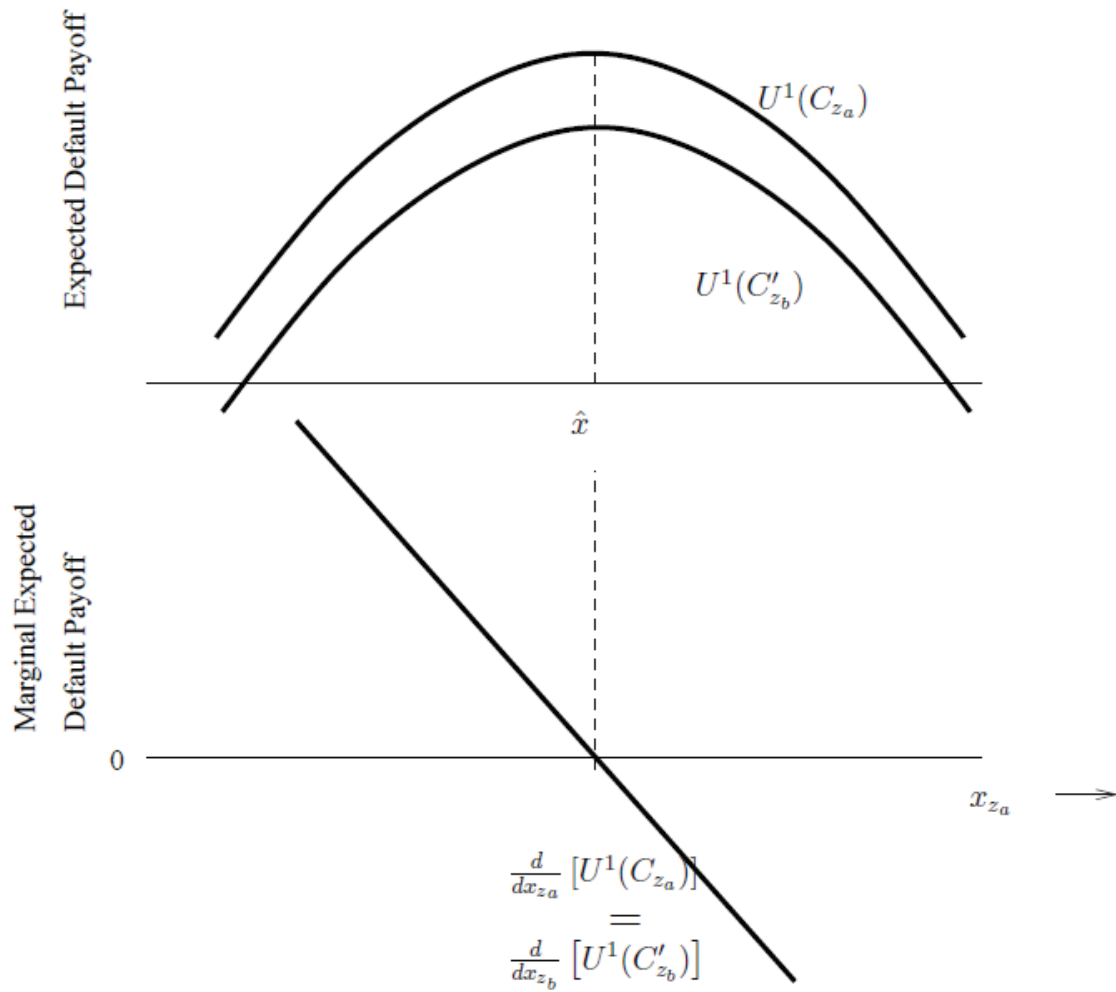


Figure 2. The polarization effect (1st derivative of U^1) is independent of the replaced member, \mathbf{a} , and the coalition choice, \mathbf{C} .

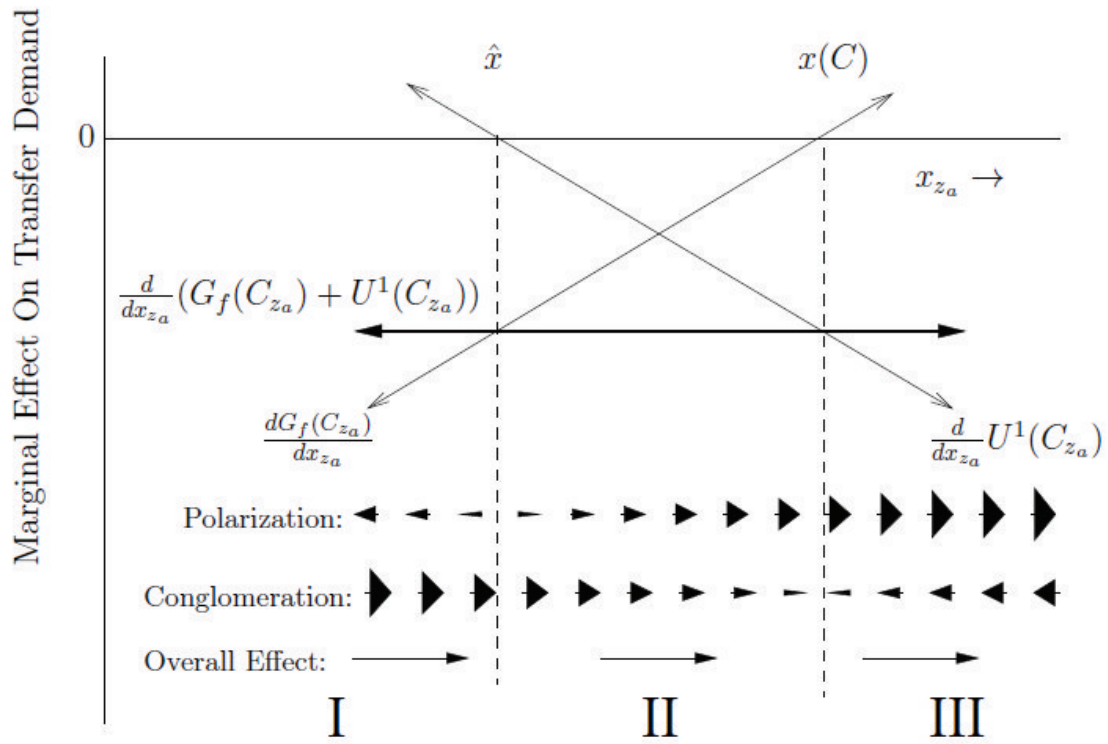


Figure 3. Decomposition of the marginal effect on transfer demand of increasing a right-leaning coalition member's ideal point.

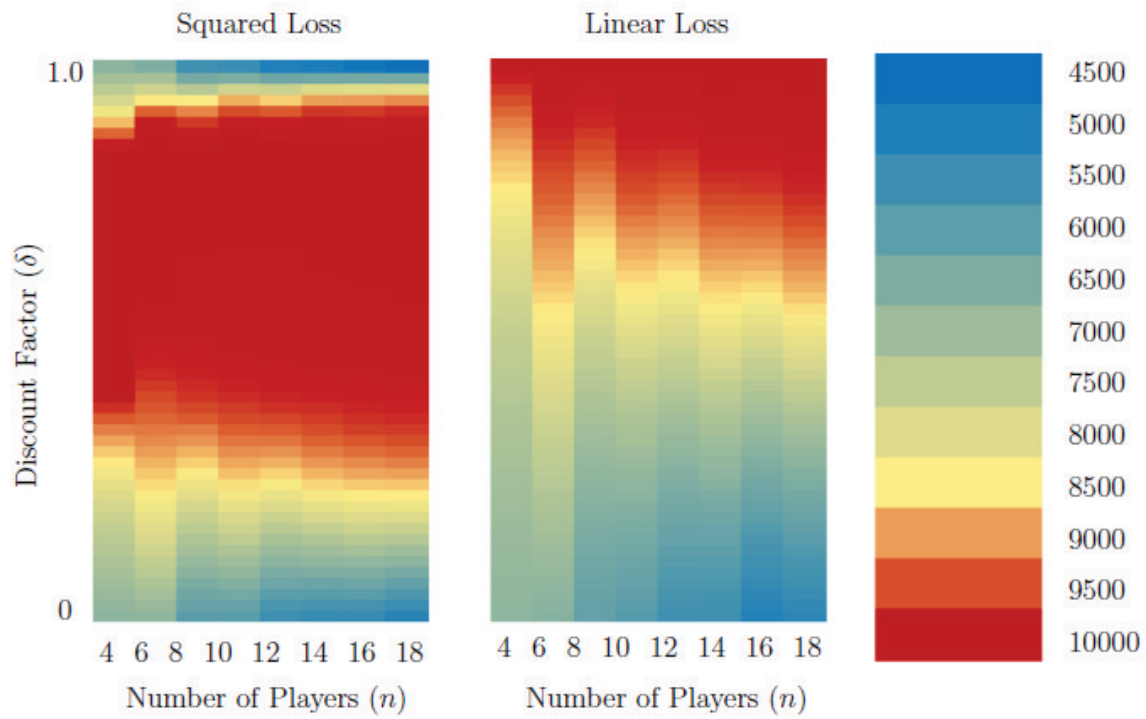


Figure 4. Frequency of Polar Coalition Equilibria under Squared and Linear Loss Functions. The color of each cell represents the number of trials out of 10,000 random symmetric ideal point configurations for which there was a polar coalition equilibrium.