

Can Democracy Cope with Extreme Views?*

Hans Gersbach
KOF Swiss Economic Institute
at ETH Zurich and CEPR
8092 Zurich, Switzerland
hgersbach@ethz.ch

Maik T. Schneider
Department of Economics
at University of Graz
Universitaetsstr. 15
8010 Graz, Austria
maik.schneider@uni-graz.at

Oriol Tejada
School of Business and Economics
at University of Barcelona
Avinguda Diagonal, 690–696
08034 Barcelona, Spain
oriol.tejada@ub.edu

This Version: July 11, 2025

Abstract

We study how democracies can manage extremist minority views through Coalition-Preclusion Promises (CPPs), where parties credibly commit to excluding extremists from governing coalitions. We identify conditions under which CPPs improve social welfare by preventing extreme policy shifts. However, these benefits depend on parties' incentives and voters' ability to coordinate. The resulting voting game features multiple equilibria, and we extend typical equilibrium selection criteria. Our findings suggest CPPs can be an effective strategy to keep extremist parties out of government.

Keywords: coalition formation, political contracts, elections, government formation

JEL: D72, D82, H55

*We would like to thank Carlos Alos-Ferrer, Arnaud Dellis, Sebastian Fehrler and Ralph Winkler for helpful comments. We are particularly grateful to Vincent Lohmann for help on the early development of paper. All errors remain our own.

1 Introduction

Democracies around the world are increasingly coming under threat by a surge in popularity of extremist views. In several developed democracies, extremist parties have even entered governments. In others, conventional parties at the center of the political spectrum are struggling to prevent extremist policy shifts. To this end, center parties in many European parliamentary democracies have made election promises to not form a coalition government with parties at the extreme political fringes. This refusal by center parties to cooperate with populist parties at the extremist left or extremist right has been termed the political "cordon sanitaire", an expression originating in the late 1980's in Belgian politics. At that time, the centrist Belgian political parties committed to exclude the far-right Flemish nationalist Vlaams Blok party from any coalition government, even if that forced the formation of grand coalition governments between ideological rivals.¹ In the recent election campaigns in Germany the term "Brandmauer" (firewall) was used to refer to the center parties' promises not to cooperate with the far-right party AfD (Alternative for Germany). As the case of Austria after the latest election on 29 September 2024 illustrates, such coalition preclusion promises are not always honored. In their election campaign the center-right party ÖVP promised not to form a coalition government with the far-right party FPÖ, but after difficult negotiations with the other center parties reneged on its promise and tried to form a coalition government with the extremists. While "Coalition-Preclusion Promises" (CPPs) are not a new phenomenon (in an Online-Appendix we provide a number of further examples), the interest in whether such CPPs can help to prevent extremist parties from entering government has skyrocketed since the rise of the extremist political fringes and the associated threats to democratic institutions.

In this paper we examine the question of whether and how Coalition-Preclusion Promises can prevent extreme policy shifts by formulating a general two stage extensive form game encapsulating the central characteristics of the political process. By making a Coalition-Preclusion Promise conventional parties can credibly preclude entering a coalition government with an extremist party after the election. There are two conventional parties, one extreme party, and an electorate with a large number of voters. A majority of voters supports one of the conventional parties (henceforth called conventional voters) and a – possibly large – minority of citizens supports the extreme party.

¹See e.g. Axelsen (2024), Chini (2019).

Parties are interested in being part of the government as they can expect material and immaterial benefits from holding office. Some of these benefits are costly to voters and are called perks. When parties cannot or do not choose to make CPPs and voters vote sincerely, each conventional party only receives a minority of votes cast and cannot form a government on its own. Forming a grand coalition of the two conventional parties will imply that perks have to be shared equally. Hence, the conventional parties are tempted to form a coalition government with the extreme party, exchanging a policy shift in the extremist policy dimension for higher perks.

Two questions emerge from this setup. First, will conventional voters act strategically, such that one conventional party obtains a majority of votes and forms a single-party government? This would avoid the policy shift towards the extreme party. Secondly, if the answer to the first question is negative, does the possibility of committing to CPPs help to keep policies moderate, by potentially forcing conventional parties into a grand coalition? To address both questions, our two-stage game is constructed as follows: in the first stage, the two conventional parties decide whether or not to make a CPP – if this option is available – and in the second stage, voters cast their votes.

We derive conditions allowing us to characterize the subgame-perfect Nash-equilibria of this game, and obtain the following insights.

At the voting stage, voters can strategically choose to coordinate and strategically vote for one of the conventional parties. As a consequence, the political game has several subgame perfect Nash equilibria. First, without CPPs, the sincere voting equilibrium would lead to the extreme party entering a government coalition with one of the conventional parties, while voters could strategically coordinate their votes on one of the conventional parties for a single party government in the two strategic voting equilibria. Most crucial, if CPPs are available, a party will only use a CPP if it can expect to gain sufficiently in vote share. This implies that for CPPs to be an effective tool to break a sincere voting equilibrium with resulting government participation of the extreme party, voters need to be expected (off-equilibrium) to strategically coordinate their votes on the party that unilaterally makes a CPP. However, in the subgame where one party unilaterally makes a CPP, there are two voting equilibria, one where voters strategically coordinate on the party making the CPP and one where they strategically coordinate on the other party not precluding a coalition government with the extremist party. Can voters be expected to play the equilibrium coordinating votes on the party unilaterally giving the CPP?

In the main part of the paper, we examine this coordination problem extending and using several equilibrium selection criteria established in the game theoretic literature. In particular, we consider four different criteria which reflect the different ways in which coordination can break down in two-player games and extend them to our game with many voters.

Our first two criteria reflect the incentives to adhere to a specific equilibrium strategy given that other players deviate with a certain probability from the equilibrium strategy profile. The essential distinction between our criteria is whether errors in coordination are independent across players or if they are correlated. The case with correlated errors is interesting as arguably increased use of social media could lead to stronger dependence of errors. In our analysis, we consider the two polar cases of perfect dependence – that is all other players play the wrong equilibrium simultaneously – and independence. We say that one equilibrium dominates another – i.e. is selected over the other – if it can withstand a higher coordination error probability. Our third and fourth criteria reflect different extensions of the standard notion of risk dominance to our voting game. We also show how the different selection criteria relate to each other.

Our main results of this part are as follows.

When voters' errors in coordination are uncorrelated the two selection criteria predict that CPPs will be used by parties in equilibrium thereby preventing the extreme party from entering government. The key reason is that this subgame-perfect Nash equilibrium will be supported by the prediction that in the off-equilibrium subgame where only one party unilaterally makes a CPP, voters will coordinate their votes on this party. This provides the incentive for parties to deviate from the situation where no CPPs are used and consequently makes CPPs² an effective tool for moderate politics. In this equilibrium, voters vote sincerely for their preferred party leading to a grand coalition between the conventional parties implementing moderate policies and no extreme policy shifts. As we show, this is typically the welfare maximal outcome and a single party government would only be socially preferred if perks are very high relative to the political polarisation of the conventional parties. Interestingly, when CPPs are unavailable (or not credible), the two selection criteria related to uncorrelated errors by voters suggest that the equilibrium with a government including the extreme party

²Throughout the paper we make the assumption that if CPPs are available and used, they are 100% credible, treating the case of non-credible CPPs as equivalent to CPPs being unavailable or unused.

will materialize. Hence, CPPs effectively prevent participation by the extremist party in government.

In the case that voters' errors are correlated, results are less clear but suggest that CPPs may not be used by parties as in the equilibrium predicted in the subgame where one party unilaterally makes a CPP the voters will coordinate their votes on the other conventional party and hence cannot provide off-equilibrium support for the subgame perfect Nash equilibrium where both conventional parties use CPPs. However, the same selection criteria would also predict that without CPPs the extreme party would not enter government, but instead voters will coordinate their votes such that one of the conventional parties will be able to form a single party government. Hence, in this case CPPs would likely not be needed to prevent the extreme party for government.

In this way our analysis suggests that CPPs can be an effective tool for democracy to deal with extreme views in situations when they are most needed. Of course, a central condition is that CPPs are credible. In the main body of the paper, we simply compare outcomes when (credible) CPPs are available and when they are not. In section 8 we discuss how such CPPs may emerge or could be introduced.

The paper is organised as follows. Next, we discuss the relation of our paper to the literature. In section 3, we develop the political game with and without the possibility for conventional parties to credibly exclude the extremist party from a coalition government via CPPs. We then characterize the possible equilibrium outcomes and provide a welfare ranking of these in Section 4. Section 5 introduces the four equilibrium selection criteria that we use in Section 6 to predict the voting equilibria in the voting subgame and in Section 7 to characterize the expected equilibrium outcomes of the political game when CPPs are available. Section 8 discusses how CPPs can be made credible and we conclude in Section 9.

2 Relation to the Literature

Our paper touches upon several strands of the literature. First, the Political Economy literature on politicians' commitments to campaign promises. While the seminal Downsian model of political competition assumes that politicians can credibly commit to any policy in the policy space (Downs, 1957), in reality campaign promises have been reneged in a large number of instances. One remedy proposed are political contracts as

in Gersbach and Schneider (2012a,b); Gersbach et al. (2019); Gersbach (2005). While they are shown to be powerful tools to solve such problems and yielding substantial welfare improvements, there are cases with multiple equilibria where only some equilibria deliver the welfare benefits.³ This motivates the focus on equilibrium selection, thereby touching upon a second strand of the literature.

The literature on equilibrium selection in Game Theory traditionally considered 2x2 games. To select an equilibrium, Harsanyi and Selten (1988) pioneered the notion of *risk dominance* and argued that the risk dominant equilibrium is the one to be predicted as the outcome of the game. Another classical approach is due to Carlsson and van Damme (1993) showing that under a certain type of global uncertainty the risk dominant equilibrium is selected. Another large literature selects equilibria based on arguments rooted in evolutionary dynamic models, such as the stochastic models by Matsui and Matsuyama (1995); Kandori et al. (1993); Young (1993). This approach focuses on the *long-run equilibrium* or *stochastically stable equilibrium* that the dynamic process converges to. Several papers have shown that the different mentioned criteria select the same, the risk-dominant equilibrium in a 2x2 game. (Hofbauer and Sorger, 1999; Morris et al., 1995) Our paper differs from this literature in that we consider a voting (sub-)game with more than two voters. The paper closest in spirit to ours is Kim (1996), which studies n -person coordination games with multiple Pareto-rankable Nash equilibria. He considers the evolutionary dynamic approaches by Kandori et al. (1993); Matsui and Matsuyama (1995) and Foster and Young (1990) as well as the selection approaches by Harsanyi and Selten (1988) and Carlsson and van Damme (1993). The paper shows that for the class of games he considers, the five criteria are equivalent for two-person games, but not for more general games with more than two players. In particular his focus is on whether the Pareto-optimal equilibrium will be selected and he only considers the evolutionary dynamic selection criteria when discussing the voting game. Our set-up shows two significant differences: First, in contrast to the class of games Kim (1996) considers, it is not necessarily the case that a player taking a particular action is no worse off when the number of opponents taking the same action increases and second, the equilibria in the voting subgame of our political game cannot be Pareto ranked. Consequently Kim (1996)'s analysis cannot be applied directly to our setting. In the present paper, we select subgame-perfect equilibria by defining selection criteria in the spirit of Kim (1996) and tailored to our game. While our

³A discussion of the literature on binding commitments in election campaigns can be found in Gersbach et al. (2019).

extension of the risk-dominance criterion in the spirit of Harsanyi and Selten (1988) to n -voters is similar to the one in Kim (1996), we additionally consider variations of the risk-dominance extension which resemble the p -dominance criteria. An action pair is p -dominant if each action is a best response to any conjecture placing at least probability p on the other player taking his action in the pair (Morris et al., 1995). Morris et al. (1995) use the p -dominance concept in a two-player setting but with more than two strategies. Our case is the opposite, we have a binary choice but more than two players. Furthermore, we distinguish whether the errors of the opponents are correlated or not.⁴

Our paper is also related to the literature on signalling devices for coordination. Andonie and Kuzmics (2012) argue that polls in an initial stage before the election can help the coordination of voting strategies so that the majority candidate will be elected. In a political game considering elections with three candidates under plurality voting, Ekmekci (2009) argues that an outside endorser can facilitate coordination among voters who may otherwise split their votes and lead to the victory of the Condorcet loser. In our model set-up there is a similar coordination problem and coordination is triggered by coalition-preclusion promises. However, there are two strategic voting equilibria where only in one the coordination works such that parties wish to use CPPs as a coordination device.

3 The Model

We consider a model of a representative democracy with a two-dimensional policy space, three political parties and a finite number of voters.

Policy Space

A policy $p = (t, d)$ is an element of $\mathcal{P} = \mathcal{T} \times \mathcal{D}$, where $\mathcal{T} = [0, 1]$ and $\mathcal{D} = \{0, \bar{d}\}$. \mathcal{T} can be interpreted as the usual *left-right policy dimension*, e.g. the level of taxation. We assume that $\bar{d} > 0$ entails implementing a discrete shift away from the status quo,

⁴More remotely related papers on equilibrium selection are Sandholm (2001), Sawa (2014) and Blume (2003). Sandholm (2001) considers a population of players repeatedly playing an m -strategy game where players myopically adjust their strategy choice to their opponents' behaviors. In this setting, he derives the notion of a $1/k$ -dominant strategy that if played by a sufficient fraction of the population leads to the $1/k$ -dominant equilibrium in the long-run. Sawa (2014) examines a game where the agents' utilities are hit by random shocks to characterize the set of equilibria that emerges when this noise vanishes. Blume (2003) examines the noise processes used in evolutionary game theory and characterises the noise processes necessary for the known stochastic stability results to hold.

which is denoted by $0 \in \mathcal{D}$, in a dimension orthogonal to \mathcal{T} . Accordingly, we refer to policies $p \in \mathcal{P}$ with $d = 0$ as *conventional* and to policies with $d = \bar{d}$ as *extreme*.

Parties

There are three parties denoted by L , R and E . By s_j , $j \in \{L, R, E\}$, we refer to the parties' vote shares which are equal to their seat shares in parliament. Each party's political orientation is reflected by its most preferred policy $p \in \mathcal{P}$, also referred to as the party platform. We assume that $p_L = (t_L, 0)$ and $p_R = (t_R, 0)$, with $t_L < t_m = \frac{1}{2} < t_R$, are the platforms of the *conventional parties* L and R respectively. Further, we assume $t_R + t_L = 1$, so conventional parties have symmetric positions in the \mathcal{T} -policy dimension around the median $t_m = \frac{1}{2}$. The third party is the *extreme party* E with platform $p_E = (t_E, \bar{d})$. While E is extremist in dimension \mathcal{D} , we assume it is not polarised on the conventional dimension and specify $t_E = t_m$.

Parties can implement their policy platforms and obtain the full share of governmental perks when they obtain the majority of votes and hence the majority of seats in parliament. Otherwise they will have to form coalitions with another party. In the appendix of the discussion paper Gersbach et al. (2025), we provide a micro-foundation of party bargaining. If a conventional party considers forming a coalition with the extremist party E , we assume that E insists on the implementation of the extreme policy \bar{d} but would be willing to transfer (most of) their share of perks to their coalition partner. The bargaining power of the extreme party in the conventional policy dimension \mathcal{T} depends on whether only one or both of the conventional parties would be willing to form a coalition with them. In the former case we say that the extreme party has weak bargaining power resulting in policy t_{jE}^w , $j \in \{L, R\}$ and in the latter case with strong bargaining power the policy will be t_{jE}^s . It is then natural to specify

$$t_L \leq t_{LE}^w \leq t_{LE}^s \leq t_m \leq t_{RE}^s \leq t_{RE}^w \leq t_R \quad (1)$$

In a *grand coalition*, the parties L and R share perks and negotiate policy according to a Nash Bargaining solution with bargaining power proportional to the seat shares in parliament. We denote the governmental policy of a grand coalition by $p_{LR} = \{t_{LR}, 0\}$, which according to our previous assumption about the symmetry of the left- and right-wing parties platforms will be close to the median policy t_m with approximately equal seat shares of the two conventional parties. In such case, we have $t_{LE}^w \leq t_{LE}^s \leq t_{LR}$ and similarly when considering party R .

The attraction for a conventional party j to form a coalition with the extremist party

E comes from possibility of implementing a conventional policy t_{jE} that is closer to their ideal point and from additional perks of office. The price is the implementation of the extremist policy change \bar{d} . As illustrated by the real world examples in the Introduction, conventional parties often have an incentive to form such coalitions. Consequently we assume for the utilities of the conventional parties $j, k \in \{L, R\}$ that

$$V_j^j > V_{jE^w}^j > V_{jE^s}^j > V_{jk}^j > V_k^j > V_{kE^s}^j > V_{kE^w}^j . \quad (2)$$

V^j indicates the utility of conventional party j from a coalition government as indicated by the subscripts. Naturally, party j most prefers a single party government of their own party. Each conventional party would prefer to form a coalition with the extremist party (preferably with weak bargaining power by party E) to a grand coalition government.⁵ If party j cannot be part of government, it would prefer a single party government of the opposed conventional party k over an extremist policy shift to \bar{d} implemented by a coalition between parties k and E .

Voters

We denote the finite set of voters by \mathcal{N} , with cardinality $n = 2m + 1$.⁶ W.l.o.g, we let $\mathcal{N} = \{0, 1, \dots, 2m\}$. Each citizen $i \in \mathcal{N}$ is characterized by their utility over the policy space. Besides policies, voters also care about perks in office. Specifically, we assume:

$$U_i(t, d, s_G) := u(|t - t_i|) - \delta_i d - \theta s_G, \quad i \in \mathcal{N} , \quad (3)$$

where $t_i \in \mathcal{T}$ and $u : [0, 1] \rightarrow \mathbb{R}_{\leq 0}$ is strictly decreasing, (weakly) concave, and $u(0) = 0$. We further assume that $t_i < t_j$ if $i < j$, and hence we denote the median voter with respect to the policy dimension \mathcal{T} by i_m , and the corresponding median voter's preferred policy by t_m . Without loss of generality we assume $t_m = \frac{1}{2}$.⁷ The parameter $\delta_i \in \{-1, 1\}$ characterizes whether a voter is in favor of the extreme policy \bar{d} or not. The last term of the voters' utility, θs_G , where $\theta > 0$ and s_G is the share of perk-seeking politicians in government, captures the disutility associated with perks obtained by members in the government. We assume that for any given two-dimensional policy, voters prefer smaller governments to larger ones, since large governments use more resources.

⁵The assumption that the conventional parties prefer a coalition with the extremist party over a grand coalition, i.e. $V_{jE^w}^j > V_{jk}^j$, is the most favourable one for the extremist party to enter government and the most interesting one for our analysis. Otherwise if $V_j^j > V_{jk}^j > V_{jE^w}^j$, there is not much of a problem.

⁶We assume that n is odd number, which eases the presentation considerably.

⁷If this is not the case, add $\frac{1}{2} - t_m$ to all the t_i , then rescale all policies, as well as the utility function, to get back to $\mathcal{T} = [0, 1]$.

Further, there is a minority of number $2e$ of extreme voters who prefer the extreme policy shift $d = \bar{d}$, and whose ideal points regarding conventional policy \mathcal{T} are symmetrically distributed around the median $t_m = \frac{1}{2}$. These voters always vote for party E , as the extremist policy shift is more important to them than conventional policy.

Among the $n - 2e$ conventional voters, there is a number e of *swing voters* on either side of the median voter who consider to vote strategically, while the other conventional voters always vote sincerely for the conventional party whose platform policy is closest to their ideal policy. Strategic voters consider voting for the party whose policy platform is opposite to their preferences on the conventional political spectrum in order to obtain more favorable policy outcomes than if they voted for the conventional party on their side of the conventional political spectrum. We assume the strategic voters to be the e voters on both sides of the conventional spectrum with their political bliss points closest to the median policy t_m . For later use, we denote the set of strategic conventional voters by $\Omega = \{0, 1, \dots, 2e\}$, and we refer by $\Omega^L = \{i \in \Omega : t_i < t_m\}$ and $\Omega^R = \{i \in \Omega : t_i > t_m\}$ to the corresponding sets of left-wing voters and right-wing voters.⁸ Within the set of strategic voters, the ideal points most to the left, respectively right are $t_L^\Omega = \min_{i \in \Omega} t_i$ and $t_R^\Omega = \max_{i \in \Omega} t_i$. The voter distribution is symmetric, so that

$$t_L^\Omega + t_R^\Omega = 1. \quad (4)$$

To reduce the complexity of the analysis we introduce the following parameter restrictions to specify realistic and interesting relative strengths of conventional voters' preferences with regard to the extreme policy and perks.

Assumption 1 (Conventional Voter Preferences)

We assume that θ (importance of perks) is sufficiently small and \bar{d} (disutility of extremist policy) is sufficiently large such that

(i) *Conventional policy is more important than perks:*

$$\theta \left(\frac{1}{2} - \frac{1}{2n} - s_E \right) < u(|t_j^\Omega - t_{jk}|) - u(|t_j^\Omega - t_k|), \quad j, k \in \{L, R\} \quad (\text{SPC})$$

and

$$\theta \left(\frac{1}{2} - s_E \right) < \bar{u}(t_m) - \bar{u}(t_j), \quad j \in \{L, R\} \quad (\text{WSPC})$$

⁸We note that as the conventional voters are in the majority and since $t_L + t_R = 1$ with median policy $t_m = \frac{1}{2}$, the composition of the electorate is completely characterized by the pair of numbers (n, e) . The smallest possible voter number is $n = 5$ which forces $e = 1$. Then there are 2 extreme voters, 1 left-wing voter, 1 right-wing voter, and the median-voter. The next possible n would be $n = 9$, with either $e = 1$ or $e = 2$.

where $\bar{u}(t) = \sum_{i \in \mathcal{N}} u(|t - t_i|)$.

(ii) *Conventional voters want to prevent extremist policy \bar{d} :*

$$\bar{d} \geq \theta_{s_E} + u(|t_j^\Omega - t_{jE}^\nu|) - u(|t_j^\Omega - t_k|), \quad j, k \in \{L, R\}, \nu \in \{s, w\} \quad (\text{PCS})$$

(iii) *Extremist policy is worse than high perks:*

$$2\bar{d} \geq \frac{\theta}{2} \left(1 + \frac{1}{1 - 2s_E} \right) \quad (\text{WC})$$

Condition (SPC) in Assumption 1(i) expresses on the left-hand side the additional costs of perks from a grand coalition over a single-party government with simple majority and on the right-hand side the utility difference for the most left-/right-wing strategic voter between the grand-coalition policy and the policy of the single party government on the opposite side of the conventional policy spectrum. Hence, this condition says that if pivotal the most left-/right-wing strategic voter would not support a single party government of the opposite conventional party against a grand coalition to avoid the larger amount of perks with the grand coalition. Condition (WC) implies that the conventional voters welfare gain from the median policy over a single party government policy is larger than the costs of additional perks coming with the grand coalition. Comparing preferences over conventional policy and the extremist policy shift, Condition (PCS) in Assumption 1(ii) specifies that the most left-/right-wing voter would prefer a single party government of the opposite conventional party to avoid a coalition government implementing the extremist policy. Finally, Assumption 1(iii) implies that the utility loss from the extremist policy is worse for the conventional voters than the utility loss from high perks in a grand coalition.

3.1 The Political Game

In the setting described in the previous subsection, we consider the following political game:

Stage 1: Campaign with Coalition-Preclusion Promises

Stage 2: Election

Stage 3: Government formation

In the first stage, parties can make Coalition-Preclusion Promises. A party's *Coalition-Preclusion Promise (CPP)* contains a (possibly empty) set of other parties that this party voluntarily promises not to form a coalition with after the election. The CPP of party $j \in \{L, R\}$ is denoted by $C_j \in \{\emptyset, E\}$.⁹ We will associate $C_j = \emptyset$ with two cases: when the conventional party j has not promised to exclude the extreme party, or when such a promise is not credible.

In the second stage, an election to form the parliament is held in which each voter $i \in N$ casts a vote $x_i \in \{L, R, E\}$.¹⁰

In the third stage, the government formation process occurs. If a conventional party j obtains a majority of seats, it will form a single-party government to implement its policy platform $p_j = (t_j, 0)$. Otherwise, both conventional parties will try to establish a coalition government. First, we consider that if both conventional parties compete for a coalition with E , the extreme party has *strong bargaining power*, and the resulting coalition between E and one conventional party, say $j \in \{L, R\}$, will implement policy $p_{jE}^s = (t_{jE}^s, \bar{d})$. Moreover, if one of the two conventional parties has more support than the other, the party with greater support (i.e. more seats) will always be the one to form a coalition with E .¹¹ If both conventional parties have the same number of seats, we agnostically assume that each party enters into coalition with E with probability $\frac{1}{2}$. Secondly, if only one conventional party j is able to form a coalition with E , the extreme party has *weak bargaining power*, and the resulting coalition will implement

⁹Technically, we have $C_j \in \{\emptyset, \{E\}\}$. We abuse notation, and always write $C = E$, instead of $C = \{E\}$.

¹⁰For simplicity, we do not consider abstention. Recall that we consider a parliamentary democracy with proportional seat allocation, i.e. the seats in parliament are distributed among the parties according to their vote shares. To simplify the analysis we assume that the number of votes translates perfectly to the number of seats in parliament. This is equivalent to assuming that the number of seats in parliament is equal to the size of the electorate. In real parliamentary democracies many effects will distort the translation of vote shares to seats, such as rounding, or a reserved number of seats per region, etc. The model could be extended by considering a "translation function" $T : |\mathcal{N}|^3 \rightarrow \text{SP}^3$, where SP would be the size of the parliament, which would translate a set of player vote numbers for each of the three parties to a number of seats in parliament. To obtain seats in parliament, parties need a certain share s_E^* of total votes, which we assume is reached by all three parties.

¹¹Possible justifications for this assumption are, that the party with more seats can always offer an ε more than the other party, in terms of perks (even though we assumed for simplification that E does not pursue perks, in reality it certainly will care about material perks, or the gestures may be symbolic, such as control of certain key government posts, etc.), or that the electorate believes that the bigger party has a popular mandate, and voters will punish a government led by the smaller party. When we introduce the Nash equilibrium voting profiles later, as a robustness check, a footnote will consider what would happen if instead we assumed that the smaller seat share conventional party always enters a coalition with E .

policy $p_{jE}^w = (t_{jE}^w, \bar{d})$. Next to (1), we specify that

$$t_{LE}^s + t_{RE}^s = t_{LE}^w + t_{RE}^w = 1. \quad (5)$$

Third, when both conventional parties bargain over the desired policy $p_{LR} = (t_{LR}, 0)$ for implementation, t_{LR} is determined as described in item (iii) of following assumption which summarizes the outcomes of the government formation stage of the political game.

Assumption 2 (Government Formation)

If a (conventional) party j has a share $s_j \geq \frac{1}{2}$ of the votes, it will form a single-party government and $p_j = (t_j, 0)$ will be implemented. In this case, $s_G = s_j$. Otherwise, when no conventional party has a majority,

- (i) if E has strong bargaining power, it will accept the proposal of the party $j \in \{L, R\}$ with the largest seat share and p_{jE}^s will be implemented. In this case, $s_G = s_j$, as recall that party E is assumed to be irrelevant for perks. If both conventional parties have the same seat shares, each party enters into coalition with E with probability $\frac{1}{2}$.
- (ii) If E has weak bargaining power and accepts the proposal of $j \in \{L, R\}$, then p_{jE}^w will be implemented. In this case, $s_G = s_j$.
- (iii) If both conventional parties have committed themselves to not forming a coalition with the extreme party, the government size is $s_G = 1 - s_E$ and the policy

$$t_{LR} = \operatorname{argmax}_{t \in [t_L, t_R]} s_L u(t - t_L) + s_R u(t_R - t) \quad (6)$$

will be implemented. In particular, u is strictly concave and if $u(x) = -\frac{1}{2}x^2$, then $t(s_L, s_R)$ is a convex combination of t_L and t_R :

$$t_{LR} = \frac{s_L}{s_L + s_R} t_L + \frac{s_R}{s_L + s_R} t_R.$$

Knowing the outcomes at the government formation stage, we can now concentrate on the following reduced political game \mathcal{G} comprising only the first two stages:

Stage 1: Potential exclusion of extreme party from government.

Stage 2: Election.

For our analysis, a strategy profile of the CPP stage is denoted by $C = (C_L, C_R) \in \{\emptyset, E\} \times \{\emptyset, E\}$ while a strategy profile at the voting stage is represented by $\sigma = (\sigma_i)_{i \in \Omega} \in \{L, R\}^\Omega$. After Stage 2, payoffs are given according to Assumption 2. For each voter $i \in \Omega$, we let $\sigma_i = q$ denote the mixed strategy of i according to which they vote L with probability q and R with probability $1 - q$.

4 Solving the Political Game

To solve the political game by backward induction, we first find the voters' best-response function to choices of CPPs by conventional parties, and then we find the equilibria of the whole game by determining the parties' optimal choices of CPPs.

4.1 The solution to Stage 2

Assuming that conventional voters only vote strategically if it leads to a more favourable outcome and otherwise vote sincerely, it trivially follows that if (C, σ) is an equilibrium of \mathcal{G} , then σ has to be of the following types:

1. Coordinated strategic voting for L : $\sigma = \sigma^L := \overbrace{(L, \dots, L)}^{2e+1}$.
2. Sincere voting: $\sigma = \sigma^{LR} := \left(\overbrace{(L, \dots, L)}^e, \frac{1}{2}LR, \overbrace{(R, \dots, R)}^e \right)$.
3. Coordinated strategic voting for R : $\sigma = \sigma^R := \overbrace{(R, \dots, R)}^{2e+1}$.

The support for each of the conventional parties in cases 1. - 3. is depicted in Table 1. Note that in the second column with sincere voting, the median voter votes with probability $1/2$ for party L leading to seat shares $s_L = \frac{1}{2}(1 - s_E) + \frac{1}{2n}$ and $s_R = \frac{1}{2}(1 - s_E) - \frac{1}{2n}$ and with probability $1/2$ for party R leading to seat shares $s_R = \frac{1}{2}(1 - s_E) + \frac{1}{2n}$ and $s_L = \frac{1}{2}(1 - s_E) - \frac{1}{2n}$. We further note that the outcome of the government formation process after the election may differ for the same voting profile depending on the CPPs made in Stage 1.

Let \mathcal{E}^C be the set of voting profiles that may arise in equilibrium given CPPs C were made in Stage 1. We can then formulate:

	σ^L	σ^{LR}	σ^R
s_L	$\frac{1}{2} + \frac{1}{2n}$	$\frac{1}{2}(1 - s_E) \pm \frac{1}{2n}$	$\frac{1}{2} - s_E - \frac{1}{2n}$
s_R	$\frac{1}{2} - s_E - \frac{1}{2n}$	$\frac{1}{2}(1 - s_E) \mp \frac{1}{2n}$	$\frac{1}{2} + \frac{1}{2n}$

Table 1: Conventional parties' support

Proposition 1 (Nash Equilibria of Voting Subgame)

In any equilibrium (in pure strategies, except for the median voter) of the subgame of \mathcal{G} starting at Stage 2 after $C \in \{\emptyset, E\} \times \{\emptyset, E\}$ was chosen, we have:

- (a) If $C = (\emptyset, \emptyset)$, $\mathcal{E}^C = \{\sigma^L, \sigma^{LR}, \sigma^R\}$.
- (b) If $C = (E, \emptyset)$, $\mathcal{E}^C = \{\sigma^L, \sigma^R\}$.
- (c) If $C = (\emptyset, E)$, $\mathcal{E}^C = \{\sigma^L, \sigma^R\}$.
- (d) If $C = (E, E)$, $\mathcal{E}^C = \{\sigma^{LR}\}$

The proof of Proposition 1 is provided in the Online-Appendix. The economic intuition for the equilibria with strategic coordinated voting σ^L and σ^R is that a deviation by a voter, e.g. voting for party R strategically, would lead to a coalition government between party R and extremist party E , which is a worse outcome than a single party government of R (and vice versa for the equilibrium σ^L). Sincere voting σ^{LR} is an equilibrium as a voter's deviation would lead to a coalition government of their less preferred conventional party with the extremist party E for sure, while otherwise there is a probability of at least $1/2$ that their preferred conventional party will form a government with E . Interestingly, with unilateral CPPs, there is no sincere voting equilibrium, as Proposition 1 (b) and (c) shows. In this case, e.g. with $C = (E, \emptyset)$, sincere voting would lead to a government between R and E . As according to Assumption 2 on the government formation process, the amount of perks depends on the governmental seat share, both types of conventional voters $i \in \Omega^L$ and $i \in \Omega^R$ have an incentive to deviate to vote for party L to reduce the seat share for R . By contrast, the strategic coordinated voting equilibria σ^L, σ^R are sustained by the fact that a deviation leads to the worse outcome of a government coalition RE instead of a single-party government by a conventional party. Finally, when both conventional parties precluded to form a government coalition with E , there is no need for strategic coordinated voting and we obtain a unique sincere voting equilibrium leading to a grand coalition

government among conventional parties.¹² Proposition 1 thus reveals that—except when both conventional parties have credibly excluded the extreme party holds—we encounter multiple equilibria. The equilibrium election results directly translate into possible government outcomes as summarized in the next corollary:

Corollary 1

- (a) If $C = (\emptyset, \emptyset)$, the following equilibrium outcomes are possible: (σ^L, L) , (σ^{LR}, LE) , (σ^{LR}, RE) , or (σ^R, R) .
- (b) If $C = (E, \emptyset)$, the equilibrium outcomes (σ^L, L) or (σ^R, R) are possible.
- (c) If $C = (\emptyset, E)$, the equilibrium outcomes (σ^L, L) or (σ^R, R) are possible.
- (d) If $C = (E, E)$, the equilibrium outcome will be (σ^{LR}, LR) .

We observe that when no CPPs are made, equilibria exist in which the extreme party enters government.¹³

4.2 Welfare Ranking of Equilibria

In this subsection, we provide a utilitarian welfare ranking of the different possible equilibrium outcomes. Let

$$W(t, d, s_G) := \frac{1}{n} \sum_{i \in \mathcal{N}} U_i(t, d, s_G)$$

denote the average utilitarian welfare when a government is elected which implements policy $p = (t, d)$ and has perk-relevant parliamentary support s_G . For each $t \in \mathcal{T}$, let

$$\bar{u}(t) := \frac{1}{n} \sum_{i \in \mathcal{N}} u(|t - t_i|)$$

denote the average utility derived in the \mathcal{T} -policy dimension when t is implemented. Note that we consider here all citizens \mathcal{N} , and not only the “swing voters” Ω . We then obtain

$$W(t, d, s_G) = \bar{u}(t) - (1 - 2s_E)d - \theta s_G \tag{7}$$

¹²We note that if and (SPC) in Assumption 2 was not satisfied, there are multiple equilibria even when $C = (E, E)$ as voters would then prefer a (politically opposite) single party government with lower perks over a grand coalition.

¹³If condition (SPC) in Assumption 1 would not hold, then three types of government are possible in the equilibrium with $C = \{E, E\}$: (σ^L, L) , (σ^{LR}, LR) or (σ^R, R) .

With this definition of voter welfare, we can formulate the following ranking of equilibria.

Proposition 2 (Welfare)

- (i) *Suppose that all voters $i \in \mathcal{N}$ are symmetrically distributed around the median. If the number of voters n is large enough, a grand coalition is welfare better than a conventional single-party government:*

$$W(t_{LR}, 0, 1 - s_E) > W(t_L, 0, s_L) = W(t_R, 0, s_R).$$

- (ii) *If n is large enough, the grand-coalition is welfare better than a coalition government with party E :*

$$W(t_{LR}, 0, 1 - s_E) > W(t_{LE}^\nu, \bar{d}, s_L) = W(t_{RE}^\nu, \bar{d}, s_R), \quad \nu \in \{s, w\}.$$

irrespective of whether the extreme party has weak or strong bargaining power.

The proof of Proposition 2 can be found in the Online-Appendix.¹⁴ Proposition 2 reveals the conflict of interest between the conventional voters and the conventional parties. While voters prefer the moderate conventional policy of a grand coalition and strongly dislike the extremist policy shift, the conventional parties not only prefer an own single-party government but also a coalition with the extremist party over a grand coalition as indicated in (2). While Proposition 1 shows that both conventional parties excluding coalitions with the extremist party via CPPs can lead to the highest voter welfare, the key question of interest is whether at the first stage of the political game the parties have an incentive to commit to such CPPs. Proposition 1 also shows that unilaterally excluding E via a credible CPP can lead to a single party government by the opposed conventional party not excluding E . Hence, a conventional party will only have an incentive to exclude E in a CPP if the strategic voters coordinate their votes in its favor.¹⁵ This highlights the importance of which of the two voting equilibria in the case of unilateral CPPs will be selected. In the next section we introduce different equilibrium selection criteria for our voting game and then characterize the equilibria of the entire political game under the different criteria to see whether CPPs can be expected to lead to welfare improvements.

¹⁴Note that if opposite to Assumption 1 \neg WSPC would hold strictly, the welfare ordering would reverse as the disutility from perks would dominate the conventional policy differences of single party government on the opposite political spectrum versus a more moderate grand coalition policy.

¹⁵Generally, it is sufficient that the conventional parties' beliefs attach a sufficiently high probability on the equilibrium where voters coordinate on the party excluding E in the case of a unilateral CPP.

5 Criteria for Equilibrium Selection

In game theory the following two ways are often used to select among multiple equilibria. First, stricter conditions than that of Nash equilibrium can be imposed. One such example is trembling-hand perfection (Selten, 1975), where each agent is assumed to “tremble” and play all actions with a certain probability. Second, a (binary) dominance relation can be defined among different equilibria, such as risk-dominance (Harsanyi and Selten, 1988).¹⁶ We approach the problem of multiplicity of equilibria in the voting subgame from the latter point of view: we define different binary relations among equilibria, namely *Domination in the presence of Uncorrelated Errors*, *Domination in the presence of Correlated Errors*, *Maximum Entropy in Equilibria*, and *Maximum Entropy in Strategies*. We relate all definitions to the existing literature.

To precisely define the different criteria, let σ' , σ'' and σ be three arbitrary strategy profiles (which could be identical), and $C = (C_L, C_R) \in \{\emptyset, E\}^2$ be some CPP pure strategy profile. The following difference in expected utilities captures the preferences of voter i either for σ'_i or σ''_i depending on their beliefs about what the other voters decide.

$$\phi_i^C(\sigma'_i, \sigma''_i, \sigma_{-i}) := \mathbb{E} [U_i((\sigma'_i, \sigma_{-i})) | C] - \mathbb{E} [U_i((\sigma''_i, \sigma_{-i})) | C]. \quad (8)$$

Note that with slight abuse of notation, we write the voter’s utilities U_i as a function of the voting strategies (σ_i, σ_{-i}) rather than policy outcomes (t, d, s_G) as in (3). The equivalence follows from the fact that given CPPs C and the votes by the electorate, Assumption 2 specifies a unique policy outcome. The detailed expected utilities are provided in Lemma 2 in the Online-Appendix. If $\phi_i^C(\sigma'_i, \sigma''_i, \sigma_{-i}) \geq 0$, then, given all other players play according to strategy profile σ , it is weakly better for player i to play σ'_i than σ''_i .¹⁷ Finally for every pair of profiles in pure strategies $\sigma_{-i}, \sigma'_{-i} \in \{L, R\}^{\Omega \setminus \{i\}}$ and for each $q \in [0, 1]$, we let $q\sigma_{-i} + (1 - q)\sigma'_{-i}$ denote a profile in mixed strategies where voter $j \in \Omega \setminus \{i\}$ votes according to σ_{-i} with probability q and according to σ'_{-i} with probability $1 - q$.

¹⁶On the one hand, the advantage of this second approach over the first is that when the relation ensures that one equilibrium dominates every other equilibrium, this seems to be a reasonable prediction of the game. On the other hand, when the relation is neither complete nor transitive, a single prediction might not be selected, leaving the multiplicity problem unsolved. A way to overcome this is based on defining a measure that translates the problem of comparing equilibria to the problem of comparing two real numbers. However, this solution comes usually at a cost of imposing further assumptions like e.g. allowing for interpersonal comparisons.

¹⁷In Online-Appendix C.1 we provide further details on the specific utility differences for the different CPPs C.

Note that by definition of the Nash equilibrium we have for every $i \in \Omega$: $\phi_i^C(\sigma'_i, \sigma''_i, \sigma'_{-i}) \geq 0$, and that if $\sigma'_i = \sigma''_i$ then the definition of ϕ_i^C implies that

$$\phi_i^C(\sigma'_i, \sigma''_i, \sigma_{-i}) = 0. \quad (9)$$

Note also that by definition of ϕ_i^C in (8),

$$\phi_i^C(\sigma'_i, \sigma''_i, \sigma_{-i}) = -\phi_i^C(\sigma''_i, \sigma'_i, \sigma_{-i}). \quad (10)$$

Our paper features a “coordination” game with more than two players. Most of the refinement concepts for coordination games are usually introduced for two-player games, leaving some “room” for their generalization to games with more players. Kim (1996) analyzes a class of coordination n -player games where every agent has the same 2-action strategy set and the same utility, and there are two Nash equilibria which coincide with the strategy profiles where all players play the same strategy. Kim introduces four different criteria to support one equilibrium over the other to predict the equilibrium outcome of the game. Each of the different criteria is introduced as the extension of a 2-player game refinement/binary relation. Moreover, Kim (1996) proves that for his class of games, all criteria collapse into one, namely risk-dominance (Harsanyi and Selten, 1988), when there are only two players. Kim’s different criteria (based on different rationales) would reduce in our game to analyzing, for each player, the following function based on ϕ_i^C as defined in (8)¹⁸:

$$\psi_i^C : \{L, R\} \times S_{-i} \rightarrow \mathbb{R}, \quad \psi_i^C(x, \sigma_{-i}) := \begin{cases} \phi^C(\sigma_{-i}) & \text{if } x = L, \\ -\phi^C(\sigma_{-i}) & \text{if } x = R. \end{cases}$$

where S_{-i} is the (product) strategy set for players other than i . The different criteria assume different beliefs on what the other players will do.

Coordination trembling refers to the possibility that strategic voters may fail to coordinate on one equilibrium, even after having reached an agreement (e.g. through communication). We consider two possibilities to incorporate potential failure in coordination on one equilibrium with respect to another equilibrium in the voting game.

First, we assume that there is a (common knowledge) probability that once agents have reached an agreement to coordinate on one equilibrium, the coordination fails

¹⁸Recall that $C = (C_L, C_R) \in \{\emptyset, E\}^2$ is an arbitrary pure CPP strategy profile.

resulting in all voters switching to coordinate on the other equilibrium.¹⁹ Second, we assume that there exists a (common knowledge) probability that each of the voters (independently of other voters) deviates from the agreed equilibrium in favour of the other equilibrium.²⁰

5.1 Correlated coordination errors

We start with considering correlated errors and introduce the following definition.

Definition 1 (Immunity to correlated errors)

Given $C = (C_L, C_R) \in \{\emptyset, E\}^2$, we say that an equilibrium σ' of the (core) voting subgame is immune to correlated errors at level ε^* w.r.t. another equilibrium σ'' , if for all players $i \in \Omega$

$$(1 - \varepsilon)\phi_i^C(\sigma'_i, \sigma''_i, \sigma'_{-i}) + \varepsilon\phi_i^C(\sigma'_i, \sigma''_i, \sigma''_{-i}) \geq 0 \quad \forall \varepsilon \text{ such that } 0 \leq \varepsilon < \varepsilon^*,$$

where $0 \leq \varepsilon^* \leq 1$.²¹

For a player i that votes differently in σ' and σ'' , the above definition requires the expected value of playing their part in equilibrium σ' not be lower than the expected value of playing their part in equilibrium σ'' , given that there is a probability $1 - \varepsilon$ that all players play the agreed equilibria σ' and a probability ε that all players decide to vote according to σ'' .²² Note that for fixed σ', σ'' , $(1 - \varepsilon)\phi_i^C(\sigma'_i, \sigma''_i, \sigma'_{-i}) + \varepsilon\phi_i^C(\sigma'_i, \sigma''_i, \sigma''_{-i})$ is a decreasing linear function of ε , thus in particular continuous in ε . This implies

¹⁹A possibility would be to introduce a (Poll) Stage in the game between Stages 2 and 3 in which all voters (or a sufficiently representative share of them) are asked about their future vote. Coordinating on this “might” be easier than coordinating on a real election. In our framework, that means that n is taken to be much lower in the poll. See also Andonie and Kuzmics (2012). In this latter paper it is shown that when the size of the poll is Poisson distributed and odd, the likelihood of coordinating in one candidate (a conventional party in our case) is higher in the poll stage rather than in the election, and hence polls can be used as coordination devices. The downside of their model is that they have to impose a high degree of symmetry.

²⁰Notice that in both cases, we do not require agents to put a positive probability on every strategy in their strategy space, as is the case for ε -perfect equilibria. Rather, we only consider mixing of Nash equilibrium strategies. Thus agents do not make “mistakes”, in the sense of a “trembling-hand equilibrium”, but rather we explicitly model their failure to coordinate on one equilibrium. That means that if one agent plays the same strategy in both equilibria, his strategy remains “fixed”.

²¹Note that when $\varepsilon^* = 0$, the condition is always vacuously true, since any statement about the elements of the empty set is true. Also note that if σ' is immune to correlated errors at level ε^* w.r.t. σ'' , it is also so at any level $\varepsilon \leq \varepsilon^*$.

²²Note that if $\sigma'_i = \sigma''_i$, by equation (9), we automatically have $(1 - \varepsilon)\phi_i^C(\sigma'_i, \sigma''_i, \sigma'_{-i}) + \varepsilon\phi_i^C(\sigma'_i, \sigma''_i, \sigma''_{-i}) = 0$, $\forall \varepsilon \in [0, 1]$.

that in the definition of immunity to correlated errors above, it is equivalent to require $(1 - \varepsilon)\phi_i^C(\sigma'_i, \sigma''_i, \sigma'_{-i}) + \varepsilon\phi_i^C(\sigma'_i, \sigma''_i, \sigma''_{-i}) \geq 0 \quad \forall \varepsilon \in [0, \varepsilon^*]$ with ε^* included.²³

In Definition 1 we have introduced a binary relation between equilibria that need not be antisymmetric. In the following, we define an antisymmetric relation, which we refer to as weak dominance in correlated errors.

Definition 2 (Equilibrium dominance in correlated errors)

Given $C = (C_L, C_R) \in \{\emptyset, E\}^2$, we say that an equilibrium σ' of the (core) voting subgame (weakly) dominates an equilibrium σ'' in the presence of correlated errors for player $i \in \Omega$ if, for each $\varepsilon^* \in (0, 1)$,

$$\begin{aligned} &\sigma'' \text{ is immune to correlated errors at level } \varepsilon^* \text{ w.r.t. } \sigma' \text{ for } i \implies \\ &\sigma' \text{ is immune to correlated errors at level } \varepsilon^* \text{ w.r.t. } \sigma'' \text{ for } i. \end{aligned}$$

We use the notation $\sigma' \succ_i^{\parallel} \sigma''$ to refer to this.

Definition 2 requires that given some level of “correlated errors” ε^* , if coordination in σ' fails w.r.t coordination in σ'' , then necessarily coordination in σ'' fails w.r.t coordination in σ' at the same level ε^* . We say that an equilibrium σ' *strictly dominates an equilibrium σ'' in the presence of correlated errors for player $i \in \Omega$* if σ' weakly dominates σ'' in the presence of correlated errors for i and σ' does not weakly dominate σ'' in the presence of correlated errors for i . We denote this by $\sigma' \succ_i^{\parallel} \sigma''$.

Moreover, we extend the antisymmetry relation to equilibria independently of player identity in the following natural way. Firstly, we say that an equilibrium σ' *weakly dominates an equilibrium σ'' in the presence of correlated errors* if it does so for every player $i \in \Omega$, and we denote this by $\sigma' \succ^{\parallel} \sigma''$. Secondly, we say that an equilibrium σ' *strictly dominates an equilibrium σ'' in the presence of correlated errors* if, for every player $i \in \Omega$ such that $\sigma'_i \neq \sigma''_i$, σ' strictly dominates σ'' in the presence of correlated errors for player i . We make no requirements about those players whose strategies are identical in σ' and σ'' , as for those players $(1 - \varepsilon)\phi_i^C(\sigma'_i, \sigma''_i, \sigma'_{-i}) + \varepsilon\phi_i^C(\sigma'_i, \sigma''_i, \sigma''_{-i}) = 0$ and $(1 - \varepsilon)\phi_i^C(\sigma''_i, \sigma'_i, \sigma''_{-i}) + \varepsilon\phi_i^C(\sigma''_i, \sigma'_i, \sigma'_{-i}) = 0 \quad \forall \varepsilon \in [0, 1]$, so that both $\sigma' \succ_i^{\parallel} \sigma''$ and

²³If σ' were immune to uncorrelated errors at level ε^* w.r.t. σ'' , but $(1 - \varepsilon^*)\phi_i^C(\sigma'_i, \sigma''_i, \sigma'_{-i}) + \varepsilon^*\phi_i^C(\sigma'_i, \sigma''_i, \sigma''_{-i}) < 0$, by continuity, there would exist a small neighbourhood of ε^* where the function is still < 0 , which is in contradiction to immunity, as defined originally in Definition 1. In what follows, we usually use the closed interval $[0, \varepsilon^*]$ when working with the definition of immunity to correlated errors.

$\sigma'' \succ_i^{\parallel} \sigma'$ will always hold. As expected, we denote strict domination in the presence of correlated errors by \succ^{\parallel} .

We can then formulate.

Proposition 3

Given $C = (C_L, C_R) \in \{\emptyset, E\}^2$ and $i \in \Omega$, the following are equivalent:

- (i) σ' dominates σ'' in the presence of correlated errors for player i .
- (ii) σ' is immune to correlated errors at level $\varepsilon^* = \frac{1}{2}$ w.r.t. σ'' for player i .
- (iii) $\phi_i^C(\sigma'_i, \sigma''_i, \sigma'_{-i}) \geq \phi_i^C(\sigma''_i, \sigma'_i, \sigma''_{-i})$.
- (iv) $\frac{1}{2}\phi_i^C(\sigma'_i, \sigma''_i, \sigma'_{-i}) + \frac{1}{2}\phi_i^C(\sigma'_i, \sigma''_i, \sigma''_{-i}) \geq 0$

The proof of Proposition 3 can be found in Online-Appendix C.2. There we also show that (inequality (iii) or (iv) holds strictly) $\Leftrightarrow (\sigma' \succ_i^{\parallel} \sigma'') \Leftrightarrow (\exists \delta > 0$ such that σ' is immune to correlated errors at level $\varepsilon^* = \frac{1}{2} + \delta$ w.r.t σ'').

Proposition 3 is intuitive in the sense that the third equivalent condition requires that the loss from deviating from the optimal strategy σ'_i given that every other player plays σ'_{-i} is greater than the loss from deviating from the optimal strategy σ''_i when every other player plays σ''_{-i} . It then seems natural that σ' is more “robust” or “stable” than σ'' , which is what our selection criterion aims to capture.

5.2 Uncorrelated coordination errors

Next, we focus on the case with uncorrelated errors and start again by defining immunity of an equilibrium to this type of error.

Definition 3 (Immunity to uncorrelated errors)

Given $C = (C_L, C_R) \in \{\emptyset, E\}^2$, we say that an equilibrium σ' of the (core) voting subgame is immune to uncorrelated errors at level ε^* w.r.t. another equilibrium σ'' , if for all players $i \in \Omega$

$$\phi_i^C(\sigma'_i, \sigma''_i, (1 - \varepsilon)\sigma'_{-i} + \varepsilon\sigma''_{-i}) \geq 0 \quad \forall \varepsilon \text{ such that } 0 \leq \varepsilon < \varepsilon^*,$$

where $0 \leq \varepsilon^* \leq 1$.²⁴

²⁴Note that when $\varepsilon^* = 0$, the condition is always vacuously true, since any statement about the elements of the empty set is true. Also note that if σ' is immune to uncorrelated errors at level ε^* w.r.t. σ'' , it is also so at any level $\varepsilon \leq \varepsilon^*$.

For a player i that votes differently in σ' and σ'' , the above definition requires that the expected value of playing their part in equilibrium σ' not be lower than the expected value of playing their part in equilibrium σ'' , given that there is a probability $1 - \varepsilon$ that each player plays according to the agreed equilibrium σ' and a probability ε that they do not and vote according to σ'' . Immunity at a level ε^* then means that if player i assumes that the errors only occur with probability at most ε^* , it is still worth playing σ' . If the voting game involves only two players, Definitions 1 and 3 are equivalent. In the case of the function $\phi_i^C(\sigma'_i, \sigma''_i, (1 - \varepsilon)\sigma'_{-i} + \varepsilon\sigma''_{-i})$ defined above, it is not as clear as in the correlated case, whether it is in fact continuous in ε for fixed σ' , σ'' . In Online-Appendix C.1 we show the continuity of $\phi_i^C(\sigma'_i, \sigma''_i, (1 - \varepsilon)\sigma'_{-i} + \varepsilon\sigma''_{-i})$ in ε , which allows us to conclude that $\phi_i^C(\sigma'_i, \sigma''_i, (1 - \varepsilon)\sigma'_{-i} + \varepsilon\sigma''_{-i}) \geq 0 \quad \forall \varepsilon \in [0, \varepsilon^*)$ is equivalent to $\phi_i^C(\sigma'_i, \sigma''_i, (1 - \varepsilon)\sigma'_{-i} + \varepsilon\sigma''_{-i}) \geq 0 \quad \forall \varepsilon \in [0, \varepsilon^*]$. In what follows we therefore use the closed interval $[0, \varepsilon^*]$.

Using Definition 3, we can now define equilibrium dominance in uncorrelated errors.

Definition 4 (Equilibrium dominance in uncorrelated errors)

Given $C = (C_L, C_R) \in \{\emptyset, E\}^2$, we say that an equilibrium σ' of the (core) voting subgame (weakly) dominates an equilibrium σ'' in the presence of uncorrelated errors for player $i \in \Omega$ if, for each $\varepsilon^* \in (0, 1)$,

$$\begin{aligned} &\sigma'' \text{ is immune to uncorrelated errors at level } \varepsilon^* \text{ w.r.t. } \sigma' \text{ for } i \implies \\ &\sigma' \text{ is immune to uncorrelated errors at level } \varepsilon^* \text{ w.r.t. } \sigma'' \text{ for } i. \end{aligned}$$

We use the notation $\sigma' \succsim_i^\perp \sigma''$ to refer to this.

As in the previous case with correlated errors, we say that an equilibrium σ' *strictly dominates* an equilibrium σ'' in the presence of uncorrelated errors for player $i \in \Omega$ if σ' weakly dominates σ'' in the presence of uncorrelated errors for i and σ'' does not weakly dominate σ' in the presence of uncorrelated errors for i . The extension to domination independently of player identity is analagous to the case of correlated errors. We use the notation \succsim^\perp for the weak partial ordering and \succ^\perp for the strict partial ordering “in the presence of uncorrelated errors”.

Unfortunately, there is no counterpart to Proposition 3 for the uncorrelated case. The reason is that even for relatively simple games, such as the core voting game considered here, the function $\phi_i^C(\sigma'_i, \sigma''_i, (1 - \varepsilon)\sigma'_{-i} + \varepsilon\sigma''_{-i})$ is not monotonic in ε . This is

intuitive, since by Lemma 3 the function is a polynomial of order $n - 1$. While one of the directions of Proposition 3 always holds, the following Proposition states that if $\phi_i^C(\sigma'_i, \sigma''_i, (1 - \varepsilon)\sigma'_{-i} + \varepsilon\sigma''_{-i})$ satisfies a “near-monotonicity” property, then the other direction holds as well.

Proposition 4

Let $C = (C_L, C_R) \in \{\emptyset, E\}^2$, $i \in \Omega$, and $f_i(\varepsilon) := \phi_i^C(\sigma'_i, \sigma''_i, (1 - \varepsilon)\sigma'_{-i} + \varepsilon\sigma''_{-i})$.

(i) The following implication always holds:

$$\begin{aligned} & \sigma' \text{ is immune to uncorrelated errors at level } \varepsilon^* = \frac{1}{2} \text{ w.r.t. } \sigma'' \text{ for player } i \\ \implies & \sigma' \text{ dominates } \sigma'' \text{ in the presence of uncorrelated errors for player } i. \end{aligned}$$

Moreover, if immunity is at level $\frac{1}{2} + \delta$ for some $\delta > 0$, then the domination is strict.

(ii) Suppose that the following “near-monotonicity” property holds: $f_i(\varepsilon) \geq f_i(\frac{1}{2}) \quad \forall \varepsilon \in [0, \frac{1}{2}]$, and $f_i(\varepsilon) \leq f_i(\frac{1}{2}) \quad \forall \varepsilon \in [\frac{1}{2}, 1]$. Then the reverse implication also holds:

$$\begin{aligned} & \sigma' \text{ dominates } \sigma'' \text{ in the presence of uncorrelated errors for player } i \\ \implies & \sigma' \text{ is immune to uncorrelated errors at level } \varepsilon^* = \frac{1}{2} \text{ w.r.t. } \sigma'' \text{ for player } i. \end{aligned}$$

Moreover, if $\sigma' \succsim_i^\perp \sigma''$ then $\exists \delta > 0$ such that σ' is immune to uncorrelated errors at level $\frac{1}{2} + \delta$ w.r.t. σ'' for i .

(iii) If $f_i(\varepsilon) + f_i(1 - \varepsilon) \geq 0 \quad \forall \varepsilon \in [0, 1]$, then σ' dominates σ'' in the presence of uncorrelated errors for player i . If $f_i(\varepsilon) + f_i(1 - \varepsilon) > 0$, domination is strict.

The proof of Proposition 4 can be found in Online-Appendix C.3. Note that item (iii) of Proposition 4 provides a different sufficient condition for domination in the presence of uncorrelated errors. In Online-Appendix C.4 we provide a Lemma stating that without further conditions, immunity at level $\varepsilon^* = \frac{1}{2}$ is in fact not necessary for $\sigma' \succsim_i^\perp \sigma''$. In that sense it completes Proposition 4. In Online-Appendix C.4 we also provide an example from our own (core) voting game.

5.3 Entropy in equilibria

Our approach in this section builds on the concept of risk-dominance (Harsanyi and Selten, 1988). For coordination games with two players $N = \{1, 2\}$, strategy set

$S_1 = S_2 = \{A, B\}$ and utilities satisfying $u_1(A, A) > u_1(B, A)$, $u_1(B, B) > u_1(A, B)$, $u_2(A, A) > u_2(A, B)$ and $u_2(B, B) > u_2(B, A)$ we say that (A, A) *risk-dominates* (B, B) if and only if the two equivalent conditions hold:

- (a) $u_1(A, A) - u_1(B, A) > u_1(B, B) - u_1(A, B)$ and $u_2(A, A) - u_2(A, B) > u_2(B, B) - u_2(B, A)$.
- (b) $u_1(A, \frac{1}{2}A + \frac{1}{2}B) > u_1(B, \frac{1}{2}A + \frac{1}{2}B)$ and $u_2(\frac{1}{2}A + \frac{1}{2}B, A) > u_2(\frac{1}{2}A + \frac{1}{2}B, B)$.

Condition (a) requires that the loss for each player of deviating from the equilibrium (A, A) be larger than the loss of deviating from the equilibrium (B, B) , whereas (b) requires the expected value for each player of A be larger than B , provided that both players are completely uncertain about what the other player will do. Notice that there are games in which no equilibrium (even if they are different) risk-dominates the other, for the incentives may be shifted from one player to another.²⁵

Next, we present several extensions of the above notion.

Definition 5 (Dominance in entropy in equilibria)

We say that an equilibrium of the (core) voting subgame σ' (weakly) dominates another equilibrium σ'' according to maximum entropy in equilibria, if for all players $i \in \Omega$,

$$\frac{1}{2}\phi_i^C(\sigma'_i, \sigma''_i, \sigma'_{-i}) + \frac{1}{2}\phi_i^C(\sigma'_i, \sigma''_i, \sigma''_{-i}) \geq 0.$$

Strict domination for player $i \in \Omega$ is defined naturally with a strict inequality. We say that the domination is strict if the above inequality is strict for all players $i \in \Omega$ such that $\sigma'_i \neq \sigma''_i$.

The above definition assumes that every voter i believes that both equilibria will occur with probability $\frac{1}{2}$. By Lemma 3.(iv), it follows that σ' dominates σ'' according to maximum entropy in equilibria for player $i \in \Omega$ if and only if σ' dominates σ'' in the presence of correlated errors. Moreover, the domination is strict according to maximum entropy in equilibria if and only if the domination in the presence of correlated errors is strict.

²⁵To take that into account, a weighted measure could be introduced.

5.4 Entropy in strategies

Focusing the entropy in strategies instead of in equilibria, we present two versions of definition.

Definition 6 (Dominance in entropy in strategies - version 1)

We say that an equilibrium of the (core) voting subgame σ' weakly dominates another equilibrium σ'' according to maximum entropy in strategies (version 1), if for all players $i \in \Omega$

$$\phi_i^C \left(\sigma'_i, \sigma''_i, \frac{1}{2}\sigma'_{-i} + \frac{1}{2}\sigma''_{-i} \right) \geq 0.$$

As in the maximum entropy definition, we extend this definition to strict domination for player i , and then to domination independently of player identity.

The above definition assumes that each voter believes that all other voters will, independently of each other, choose to play each equilibrium strategy with probability $\frac{1}{2}$. By Proposition 4, it follows that if $\phi_i^C \left(\sigma'_i, \sigma''_i, (1 - \varepsilon)\sigma'_{-i} + \varepsilon\sigma''_{-i} \right)$ satisfies the “near-monotonicity” property described therein, domination according to maximum entropy in strategies (version 1) is equivalent to domination in the presence of uncorrelated errors, including the respective strict cases. However, as Lemma 4 shows, the “near-monotonicity” property is by no-means assured. Thus, we expect that examples of games and equilibria exist where domination according to maximum entropy in strategies (version 1) is truly different from domination in the presence of uncorrelated errors. In our voting game, no such examples exist however.

Definition 7 (Dominance in entropy in strategies - version 2)

We say that an equilibrium of the (core) voting subgame σ' dominates another equilibrium σ'' according to maximum entropy in strategies (version 2), if for all players $i \in \Omega$

$$\phi_i^C \left(\sigma'_i, \sigma''_i, \frac{1}{2}\sigma^L_{-i} + \frac{1}{2}\sigma^R_{-i} \right) \geq 0.$$

As before, we extend the definition to strict domination, and independently of player identity.

The above definition assumes that each voter believes that every other voter votes for either party with probability $\frac{1}{2}$.

6 Equilibrium Selection for the Voting Subgame

6.1 Voting Equilibria without Credible Exclusion Promises

Using the equilibrium selection criteria introduced in the previous section, we start by discussing the selection of equilibria without CPPs, $C = (\emptyset, \emptyset)$. According to Proposition 1, we have $\mathcal{E}^C = \{\sigma^L, \sigma^{LR}, \sigma^R\}$. We make pairwise comparisons among these equilibria to find out which is preferred according to the different criteria. For some comparisons the size of the electorate n matters. In these cases, when we say that a statement holds “if n and e are large enough”, we mean that given an arbitrary parliamentary seat-share $\frac{1}{2} > \tilde{s}_E > s_E^*$ for the extreme party, there exists an $N \in \mathbb{N}$ such that for all $n > N$ and $e \geq e(n) := \lfloor \frac{\tilde{s}_E n}{2} \rfloor$, the statement holds.²⁶ That is, we make the asymptotic assumption that e increases at least as fast as $e(n)$, such that the extreme party seat share is always below but close to the given parliamentary seat share \tilde{s}_E . We can now formulate:

Proposition 5 (Selection of Voting Equilibria without CPPs)

- (i) *Equilibria with voter coordination on one single conventional party σ^L and σ^R strictly dominate the sincere voting equilibrium σ^{LR} in the presence of correlated errors, if $n \geq 9$ and $e \geq 2$. I.e. $\sigma^L, \sigma^R \succ^{\parallel} \sigma^{LR}$, if $n \geq 9$ and $e \geq 2$.*
- (ii) *If n and e are large enough, the sincere voting equilibrium σ^{LR} is immune to uncorrelated errors at level $\frac{1}{2}$ with respect to both coordinated voting equilibria σ^L and σ^R . In particular, σ^{LR} strictly dominates σ^L and σ^R in the presence of uncorrelated errors, i.e. $\sigma^{LR} \succ^{\perp} \sigma^L, \sigma^R$.*
- (iii) *The coordinated voting equilibria σ^L and σ^R strictly dominate the sincere voting equilibrium σ^{LR} according to maximum entropy in equilibria, if $n \geq 9$ and $e \geq 2$.*
- (iv) *The sincere voting equilibrium σ^{LR} strictly dominates the coordinated voting equilibria σ^L and σ^R according to maximum entropy in strategies (version 1), if n and e are large enough.*

The proof of Proposition 5 can be found in Appendix A.1.²⁷ The proposition states

²⁶The seat share \tilde{s}_E is arbitrary but satisfying the assumptions made in section 3.1, i.e. that the extremist party enters parliament $\tilde{s}_E > s_E^*$ and is not able to form a single party government $\tilde{s}_E < \frac{1}{2}$.

²⁷We further note that in Proposition 5 (ii), given $\bar{t} \in (\frac{1}{2}, 1]$ (in addition to all the usual model parameters), if n and e are large enough, the ‘entropy in strategies (version 1)’ and ‘uncorrelated errors’ criteria are equivalent for the set of players $\Omega^L \cup \{i \in \Omega^R : t_i \geq \bar{t}\}$, in the case of comparison of σ^{LR} to σ^L , and for the set of players $\Omega^R \cup \{i \in \Omega^L : t_i \leq 1 - \bar{t}\}$, in the case of comparison to σ^R .

that without CPPs, we expect the sincere voting equilibrium σ^{LR} to materialize when errors are uncorrelated and according to maximum entropy in strategies. Without credible CPPs, the sincere voting equilibrium will then result in a government with the extremist party. By contrast, when errors are correlated or with the selection criterion of maximum entropy in equilibria, we expect one of the strategic coordinated voting equilibria to occur. In this case, a single government by a conventional party will be formed.

Intuitively, for a voting equilibrium σ' , e.g. $\sigma' = \sigma^R$ to be dominant relative to another equilibrium σ'' , e.g. $\sigma'' = \sigma^{LR}$, the minimal error probability at which the conventional voter will deviate from voting according to equilibrium σ' , e.g. a strategically R -voting individual according $\sigma' = \sigma^R$ deviating to sincerely voting L , must be larger than the minimal error probability at which the conventional voter will deviate from voting according to equilibrium σ'' , e.g. a sincerely L -voting individual according to $\sigma'' = \sigma^{LR}$ deviating to strategically voting R .

In the case with correlated errors, the strategically R voting individual in equilibrium $\sigma' = \sigma^R$ is pivotal for a single party government of R with probability $1 - \varepsilon$, while with probability ε the other voters deviate towards their strategies in $\sigma'' = \sigma^{LR}$. The expected loss from deviating from σ' is the utility difference between a single party R -government and an RE coalition government, weighted with probability $1 - \varepsilon$, plus the utility difference between a RE government and a RE and LE government with probability one half each, weighted with probability ε . As the former utility difference between the single party government and the coalition with the extremists is larger than the latter utility difference between the different coalition government options including E , there exist values of ε where the strategic voters would stick to voting strategically according to σ^R with correlated errors towards σ^{LR} , but voters would deviate at that error level ε from sincere voting σ^{LR} with correlated errors towards σ^R . Thus, σ^R dominates equilibrium σ^{LR} when errors are correlated, and analogously for σ^R versus σ^{LR} .

This is different for uncorrelated errors. When every voter has a probability of ε to deviate from the intended strategy, the probability of the strategically R -voting individual to be pivotal for a single party R government in equilibrium σ^R is only $(1 - \varepsilon)^{2e-1}$. With the complementary probability the government will be some coalition between one of the conventional parties and E , in which case sincerely voting for L would be preferable. Even though, the utility difference between a single party government by

conventional party R and some coalition government with E is larger than the utility difference between different forms of coalitions between a conventional party and E , for a very large ranges of ε it is much more likely to be pivotal for some form of coalition government with E , either for LE vs. RE or the amount of perks in such coalitions, than for an R -government versus an RE -government. Therefore, the minimal error probability at which strategically R -voting individuals deviate from σ^R to sincerely voting L is lower than the minimal error probability at which sincerely L -voting agents deviate from σ^{LR} to strategically voting R . Hence, in the case with uncorrelated errors, the sincere voting equilibrium σ^{LR} dominates the strategic coordinated voting equilibrium σ^R , and analogously for σ^{LR} versus σ^L .

However, as the next proposition states, due to the symmetry of our set-up, our selection criteria are unable to choose between the two strategic coordinated voting equilibria.

Proposition 6 (Non-Comparability of Coordinated Voting Equilibria)

The coordinated voting equilibria σ^L and σ^R are not comparable (neither dominates the other) according to the orderings given by the correlated errors criterion, the uncorrelated errors criterion, the maximum entropy in equilibria criterion, and maximum entropy in strategies criterion.

The proof of Proposition 6 can be found in Online-Appendix C.5. As we show in the next section, credible unilateral CPPs break this symmetry between the coordinated voting equilibria, which will have important implications for the conventional parties' incentives to commit to CPPs in the first place.

6.2 Voting Equilibria with Credible Exclusion Promises

We now consider the case where during campaigns, the conventional parties can commit to not forming a coalition with the extreme party to govern the country. In Section 8, we will detail which type of commitment devices might be available. Let $C = (E, \emptyset)$. According to Proposition 1, we have $\mathcal{E}^C = \{\sigma^L, \sigma^R\}$. Comparing these two equilibria according to our selection criteria, we can formulate the following proposition. There, the meaning of “if n and e large enough” is identical to what it was in Proposition 5, implying that e and n grow at an approximately constant ratio.

Proposition 7 (Selection of Voting Equilibria with Unilateral CPPs)

(i) *The strategic coordinated voting equilibria σ^R and σ^L are not comparable (neither*

dominates the other) in the presence of correlated errors, unless n is large enough (irrespective of e) and the voters disutility from perks θ is sufficiently large. In the latter case σ^R strictly dominates σ^L in the presence of correlated errors.

- (ii) If n and e are large enough, σ^L is immune to uncorrelated errors at level $\frac{1}{2}$ w.r.t. σ^R , and in particular σ^L strictly dominates σ^R in the presence of uncorrelated errors, i.e. $\sigma^L \succ^\perp \sigma^R$.
- (iii) The strategic coordinated voting equilibria σ^R and σ^L are not comparable (neither dominates the other) according to maximum entropy in equilibria, unless n is large enough (irrespective of e) and the voters disutility from perks θ is sufficiently large. In the latter case σ^R strictly dominates σ^L according to maximum entropy in equilibria.
- (iv) If n and e are large enough, σ^L strictly dominates σ^R according to maximum entropy in strategies (version 1 and version 2, as these are equivalent here).

The proof of Proposition 7 can be found in Appendix A.2. Note that by symmetry, the exact same statements, *mutatis mutandis* will hold when $C = (\emptyset, E)$. In Proposition 7 (ii) and (iv), parameter θ indicating the voters disutility from perks needs to satisfy the following condition:

Large Perks Condition (LPC): $s_E \theta > u(|t_j - t_j^\Omega|) - u(|t_k - t_j^\Omega|)$.

The key insight of Proposition 7 is that unilateral CPPs by one of the conventional parties can break the non-comparability of the strategic coordinated voting equilibria σ^L and σ^R . In particular, in the case of uncorrelated errors and according to the criterion of maximum entropy in strategies, we now expect voter coordination on the party that excludes a coalition with the extremist party E .

For an intuition of the result, consider the situation where party L precludes E from a coalition while party R does not. Then a voter deviating from σ^R or σ^L will lead to a RE coalition government. Therefore in the correlated errors case, when comparing the two equilibria with respect to their expected losses from deviating from the agreed upon equilibrium, the difference in the resulting RE governments when deviating from σ^L versus deviating from σ^R is the governmental seat share s_G and hence the amount of perks. Therefore, only if the voters' disutility from perks θ is sufficiently large, will we observe equilibrium dominance in the correlated errors case.²⁸ It may be counterintu-

²⁸However, note that condition (LPC) violates condition (SPC) in Assumption 1.

itive that the party that did not exclude E will benefit from the other party excluding the extremist party. However, a deviation from σ^R will lead to an RE coalition government with substantially higher perks than a deviation from σ^L , and therefore there exist correlated error probabilities where strategic voters would still vote for R in a σ^R equilibrium, but strategic voters would deviate from L in a σ^L equilibrium.

With uncorrelated errors this changes. Then, the larger difference in utilities from policy t_L in the single party government versus t_{RE} in the coalition government relative to that between t_R and t_{RE} is the dominant factor. Due to this the minimum error probability where strategically R -voting individuals deviate from σ^R is lower than where strategically L -voting individuals deviate from σ^L , which implies equilibrium dominance of σ^L . Again this intuition is valid vice versa for the situation where $C = \{\emptyset, E\}$.

7 Solving the Political Game when CPPs are Credible

We next investigate whether promises to exclude E are made in equilibrium, assuming that CPPs are available, that is, we now assume that the parties are able to endogenously select which CPP they wish to make. In doing so, the conventional parties take into account which voting equilibria will result in the subgame after having decided on their CPPs. Using our previous results, without exclusion promises of E , $C = (\emptyset, \emptyset)$, we expect coordinated voting equilibria leading to a single party government according to our selection criteria when errors are correlated. By contrast, our selection criteria with uncorrelated errors predict a sincere voting equilibrium leading to a coalition government with E , that is a government of (LE) or (RE) each with probability $\frac{1}{2}$. According to Proposition 7, with unilateral CPPs we predict a coordinated voting equilibrium favoring the party that excludes E when using the section criteria when errors are uncorrelated, while voter coordination on the party that does not preclude E may occur with correlated errors. Finally when both parties exclude E , $C = (E, E)$, the unique equilibrium is the sincere voting equilibrium leading to a grand coalition. The following table depicts the game at Stage 1 where parties choose their CPPs when the equilibrium selection criteria with uncorrelated errors are used.

From the party preferences (2), it follows immediately that $C = (E, E)$ is the only

		Party R	
		\emptyset	E
Party L	\emptyset	$\frac{1}{2}(V_{LE^s}^L + V_{RE^s}^L)$, $\frac{1}{2}(V_{LE^s}^R + V_{RE^s}^R)$	V_R^L , V_R^R
	E	V_L^L , V_L^R	V_{LR}^L , V_{LR}^R

Table 2: Conventional parties' contract choice game

equilibrium of the game.²⁹ In this situation, CPPs are most effective in preventing extreme parties from entering government and ensuring moderate policy. Our previous results in Propositions 5 and 7 regarding equilibrium selection in the different subgames after CPPs have been chosen, indicate when the above equilibrium of the entire political game will be predicted. We summarize our main result in the following proposition:

Proposition 8 (Equilibria in Political Game with CPPs)

- (i) *In an equilibrium the political game, either both conventional parties make CPPs $C = \{E, E\}$ or no CPPs are made $C = \{\emptyset, \emptyset\}$.*
- (ii) *If n and e are large enough, and equilibria in the voting subgames, are predicted by the selection criteria “Domination in the presence of uncorrelated errors” or “maximum entropy in strategies”. Then the availability of CPPs prevents extreme parties from entering government.*

That is, the predicted unique equilibrium of the political game is: $C = \{E, E\}$, voters vote sincerely according to σ^{LR} leading to a grand coalition government of L and R .

- (iii) *The selection criteria “Domination in the presence of correlated errors” or “maximum entropy in equilibria” provide no clear prediction whether $C = \{E, E\}$ or $C = \{\emptyset, \emptyset\}$ will occur in equilibrium.*

However, with $C = \{\emptyset, \emptyset\}$, the selection criteria “Domination in the presence of correlated errors” or “maximum entropy in equilibria” predict strategic coordinated voting equilibria, σ^L and σ^R , leading to a single party government of either L or R .

²⁹Inspection of the choice game table 2 below shows that playing E would continue to be a strictly dominant strategy for both conventional parties, as long as there was a positive probability of either (LE) or (RE) occurring when $C = (\emptyset, \emptyset)$ (the top-left cell in the table). Hence, which equilibrium is selected when $C = (\emptyset, \emptyset)$ is not crucial for solving the full game.

Proposition 8 reveals that democracy can cope with minority extreme views when two conditions are met. First, conventional parties can credibly promise that they will not team up with the extreme party to form a government. Second, by using CPPs unilaterally the conventional parties can expect a sufficiently large gain in voter support.

The latter will be the case if the voting equilibria in the subgame where only one party uses a CPP can be predicted either by dominance in uncorrelated errors or by entropy in strategies according to Proposition 7(*ii*) and (*iv*). Interestingly, when equilibria are predicted by these two selection criteria, without the possibility of CPPs the extreme party will enter government as indicated in Proposition 5(*ii*) and (*iv*). Hence CPPs are predicted to be used when they will be needed most.

It is most realistic that voting errors in an election are largely uncorrelated.³⁰ However, we also discussed the case where errors, respectively deviations from the coordination equilibrium may be correlated. Then, Proposition 7 predicts that the party making a unilateral CPP could not expect higher but possibly lower voter support. Hence, parties would not use CPPs.³¹ It is interesting, however, that according to the same selection criteria Proposition 5 predicts that the extremist party will not enter government but rather a single party government by one of the conventional parties will be formed in the case when no CPPs are made.³² Consequently, when the selection criteria assuming correlated errors are applied, CPPs are not necessary to prevent the extremist party from entering government.

In summary, the expected equilibrium outcome of the political game is a single-party government according to the selection criteria with correlated errors, while according to the selection criteria with uncorrelated errors, the predicted equilibrium outcome

³⁰Individuals making coordination errors independently seems much more realistic than dependent errors. It is more agnostic, in the sense that a correlated error where the whole population except one individual makes the same error in effect already represents a high degree of coordination. In that sense it is almost inconsistent to talk about a “coordinated error in coordination”. Uncorrelated errors are also much more realistic as n gets large and the society under consideration becomes heterogeneous. Even if there is some event or signal which might allow coordination on the “wrong” equilibrium, it is an extremely strong assumption, that not even one voter in $\Omega \setminus \{i\}$ would incorrectly read the signal, or be unaware of the signal, and not change their behaviour simultaneously with all other voters.

³¹If $C = (E, \emptyset)$, by Proposition 7, we have $\sigma^R \succ^{\parallel} \sigma^L$, which would go against our assumption that σ^L is selected. However note that this domination in the presence of correlated errors requires the PPC to hold. In that sense again, the assumption we make is more robust.

³²When $C = (\emptyset, \emptyset)$, and we use our ‘correlated errors’ equilibrium selection criterion, then it is true that both σ^L and σ^R dominate σ^{LR} , and this would go against our assumption that σ^{LR} is selected. However, in the absence of a coordination mechanism to decide which of σ^L or σ^R is selected, the prediction made by the correlated errors criterion is very unstable.

is a grand coalition. In terms of welfare, we summarize our results with the following Corollary.

Corollary 2 (Welfare Consequences of Credible CPPs)

- (i) *If n and e are large enough, and equilibria are selected according to one of the four criteria introduced in section 5, then the possibility to use (credible) CPPs is weakly welfare improving.*
- (ii) *If equilibria are selected according to “domination in the presence of uncorrelated errors” or “maximum entropy in strategies”, CPPs are strictly welfare improving.*
- (iii) *If equilibria are selected according to “domination in the presence of correlated errors” or “maximum entropy in equilibria”, CPPs are unused when PPC holds, and may or may not be used if PPC does not hold, depending on parties’ beliefs about voter coordination.*

Finally we note that the reason for the grand coalition providing higher welfare than a single-party government is the voters preferences give a relatively higher weight on political polarization of party platforms in the conventional policy dimension vis-a-vis perks. This is captured in condition WSPC in Assumption 1(i). By contrast, if perks involve very high disutility, a single-party government may deliver higher welfare than a grand coalition.

8 How to make CPPs credible

So far we have ignored the question of when CPPs are actually credible in reality, and only considered the two polar cases where CPPs are either a) non-existent, completely non-binding or irrelevant or b) absolutely binding or never reneged upon. There are traditional reasons why coalition promises are kept, mainly reputation loss. One could also envision that specific institutions be put in place to make such promises credible as e.g. the “Coalition-Preclusion Contracts” introduced in Gersbach et al. (2019).

We elaborate on both directions in this section. Most of the work on political campaigns has assumed, in the tradition of Downs (1957), that candidates can commit to policies they would implement if they succeed in taking public office. Aragonès et al. (2007) provide a detailed analysis of the credibility – or lack thereof – of campaign promises. They analyze a dynamic model in which ideologically motivated candidates

make promises during campaigns. Voters form beliefs about the policies a candidate will pursue in office, based on their promises. Voters discipline a candidate by believing a subset of their promises, as long as they have never reneged on a promise in office in the past. The value of reputation implies that a candidate’s optimal strategy consists of making promises that are, not excessively costly, believed by voters, and that they deliver if elected. These considerations can be applied to parties. This would imply that the degree to which CPPs are credible is an increasing function of the value of a party’s reputation.

A second approach to make CPPs credible is to introduce new institutions that allow parties formally to commit to such promises. Such formal commitments have been developed in Gersbach et al. (2019). In the context of a three voter model, they study how constitutional law can be used to establish “Coalition-Preclusion Contracts”. With such contracts, parties can formally exclude one party from the range of possible government coalitions. Such contracts may help to keep an extreme party from entering government.

In this paper we are agnostic about the sources of power of parties to credibly exclude extreme parties from government. We investigate how such promises may help a large pool of voters to coordinate their votes on the party making these promises.

9 Conclusion

In the recent political environment characterised by large uncertainty and crises, extremist parties are on the rise and push into governments with simple political recipes. In this paper, we examine whether the availability of (credible) Coalition-Preclusion Promises (CPPs) can prevent extremist parties from entering government. Setting up a simple political game where parties can make CPPs before an election, we identify multiple equilibria. To predict which equilibrium will be played, we extend the well-known risk dominance criterion to the political economy context and additionally use a selection criterium based on stability in the presence of unintended errors by voters.

We find that CPPs can play an effective role in preventing extreme parties entering government if errors of voters are uncorrelated. In this case, we predict that without the availability of CPPs, extreme policy shifts would be implemented. The availability of CPPs can prevent this and lead to moderate policy. While CPPs tend not to

be as effective when voters' errors are correlated, we also find that in this case it is unlikely that the extreme party will enter government without CPPs. Consequently, our analysis suggests that the introduction of (credible) CPPs can be an effective tool for democracy to deal with extreme views in situations where they are most needed.

The results of this paper suggest that future research on how to make Coalition-Preclusion Promises credible within the political process is highly valuable.

References

- AINSWORTH, P. (2016): “MLA backs event promoting teaching children creationism,” *The Irish News*, 21 September 2016. Available from: <http://www.irishnews.com/news/2016/09/21/news/mla-backs-conference-promoting-the-teaching-of-creationism-in-schools-702781>, [Accessed 24 April 2019].
- ANDONIE, C. AND C. KUZMICS (2012): “Pre-election polls as strategic coordination devices,” *Journal of Economic Behavior & Organization*, 84, 681–700.
- APOSTOLOVA, V., L. AUDICKAS, C. BAKER, A. BATE, R. CRACKNELL, N. DEMPSEY, O. HAWKINS, R. MCINNES, T. RUTHERFORD, AND E. UBEROI (2019): “General Election 2017: results and analysis, 2nd ed.” *House of Commons Library Briefing Paper*, CBP 7979, 29 January 2019.
- ARAGONÈS, E., T. PALFREY, AND A. POSTLEWAITE (2007): “Political reputations and campaign promises,” *Journal of the European Economic Association*, 5, 846–884.
- AXELSEN, J. E. (2024): “The cordon sanitaire: a social norm-based model,” *Journal of Elections, Public Opinion and Parties*, 34, 277–297.
- BARRACLOUGH, B. (2017): “NZ’s new Government: NZ First chooses Labour,” *Newshub*, 19 October 2017. Available from: <http://www.newshub.co.nz/home/election/2017/10/nz-s-new-government-nz-first-chooses-labour.html>, [Accessed 24 April 2019].
- BBC NEWS (2015): “Same-sex marriage: Proposal wins assembly majority but fails over DUP block,” *BBC*, 2 November 2015. Available from: <http://www.bbc.com/news/uk-northern-ireland-politics-34692546>, [Accessed 24 April 2019].
- (2018a): “Supreme Court rejects NI abortion law case,” *BBC*, 7 June 2018. Available from: <http://www.bbc.com/news/uk-northern-ireland-44395150>, [Accessed 24 April 2019].
- (2018b): “West Tyrone by-election candidates list,” *BBC*, 26 April 2018. Available from: <http://www.bbc.com/news/uk-northern-ireland-politics-43751656>, [Accessed 24 April 2019].
- BEAVIS, B. AND I. M. DOBBS (1990): *Optimization and Stability Theory for Economic Analysis*, New York: Cambridge University Press.

- BIRCHARD, K. (2000): “Northern Ireland resists extending abortion Act,” *The Lancet*, 356, 52.
- BLUME, L. (2003): “How noise matters,” *Games and Economic Behavior*, 44, 251–271.
- BURR, L. (2017): “The comprehensive list of Winston Peters’ bottom lines,” *Newshub*, 19 October 2017. Available from: <http://www.newshub.co.nz/home/election/2017/10/the-comprehensive-list-of-winston-peters-bottom-lines1.html>, [Accessed 24 April 2019].
- CARLSSON, H. AND E. VAN DAMME (1993): “Global games and equilibrium selection,” *Econometrica*, 61, 989–1018.
- CHINI, M. (2019): “The rule to keep the far-right out of Parliament: what is the cordon sanitaire?” *The Brussels Times* on 27 May 2019.
- DOWNS, A. (1957): “An Economic Theory of Political Action in a Democracy,” *Journal of Political Economy*, 65, 135–150.
- EKMEKCI, M. (2009): “Manipulation through political endorsements,” *Journal of Economic Theory*, 144, 1227–1248.
- ELECTORAL COMMISSION (2017): “2017 General Election - Official Result,” Available from: http://www.electionresults.govt.nz/electionresults_2017, [Accessed 24 April 2019].
- FOSTER, D. AND P. YOUNG (1990): “Stochastic Evolutionary Game Dynamics,” *Theoretical Population Biology*, 38, 219–232.
- GERSBACH, H. (2005): *Designing Democracy: Ideas for Better Rules*, Springer.
- GERSBACH, H. AND M. T. SCHNEIDER (2012a): “Tax Contracts and Elections,” *European Economic Review*, 56, 1461–1479.
- (2012b): “Tax Contracts and Government Formation,” *Mathematical Social Science*, 64, 172–193.
- GERSBACH, H., M. T. SCHNEIDER, AND O. TEJADA (2019): “Coalition preclusion contracts and moderate policies,” *Games and Economic Behavior*, 114, 28–46.

- (2025): “Can Democracy Cope with Extreme Views,” *CEPR Discussion Paper 56593-1751224357*.
- HARSANYI, J. AND R. SELTEN (1988): *A General Theory of Equilibrium Selection in Games*, MIT Press.
- HOFBAUER, J. AND G. SORGER (1999): “Perfect Foresight and Equilibrium Selection in Symmetric Potential Games,” *Journal of Economic Theory*, 85, 1–23.
- KANDORI, M., G. J. MAILATH, AND R. ROB (1993): “Learning, Mutation, and Long-run Equilibria in Games,” *Econometrica*, 61, 29–56.
- KIM, Y. (1996): “Equilibrium Selection in n-Person Coordination Games,” *Games and Economic Behavior*, 15, 203–227.
- MATSUI, A. AND K. MATSUYAMA (1995): “An Approach to Equilibrium Selection,” *Journal of Economic Theory*, 65, 415–434.
- MORRIS, S., R. ROB, AND H. S. SHIN (1995): “p-Dominance and Belief Potential,” *Econometrica*, 63, 145–157.
- NEW ZEALAND FIRST (2019): “Fifteen Fundamental Principles,” *Party Website*, Available from: www.nzfirst.org.nz/about, [Accessed 24 April 2019].
- SANDHOLM, W. H. (2001): “Almost global convergence to p-dominant equilibrium,” *International Journal of Game Theory*, 30, 107–116.
- SAWA, R. (2014): “Coalitional Stochastic Stability in Games, Networks and Markets,” *Games and Economic Behavior*, 88, 90–111.
- SELTEN, R. (1975): “Reexamination of the perfectness concept for equilibrium points in extensive games,” *International Journal of Game Theory*, 4, 25–55.
- YOUNG, H. P. (1993): “The evolution of conventions,” *Econometrica*, 61, 57–84.

Appendix

A Proofs

This Appendix provides the proofs of Propositions 5 and 7, which we deem most informative beyond the explanations in the main text for understanding our key results. All other proofs are relegated to the supplementary Online-Appendix. Also note that Lemmas 1 – 3 together with their proofs can be found in the Online-Appendix as well.

A.1 Proof of Proposition 5

(i) and (iii).

We prove that $\sigma^R \succ_{\parallel} \sigma^{LR}$ (i.e. domination is in the presence of correlated errors). According to Lemma 2.(iii), this is equivalent to showing

$$\phi_i^C(\sigma_i^R, \sigma_i^{LR}, \sigma_i^R) > \phi_i^C(\sigma_i^{LR}, \sigma_i^R, \sigma_i^{LR})$$

for all i with $\sigma_i^R \neq \sigma_i^{LR}$, that is, for all $i \in \Omega^L \cup \{m\}$.

Consider first $i \in \Omega^L$.

We calculate the distribution of $\chi_{-i}^{\sigma^R}$ and $\chi_{-i}^{\sigma^{LR}}$, in order to use Lemma 2 for the calculation of $\phi_i^C(\sigma_i^R, \sigma_i^{LR}, \sigma_i^R)$ and $\phi_i^C(\sigma_i^{LR}, \sigma_i^R, \sigma_i^{LR})$. We have

$$p(\chi_{-i}^{\sigma^R} = j) = \begin{cases} 1 & \text{if } j = 0, \\ 0 & \text{otherwise.} \end{cases}$$

and

$$p(\chi_{-i}^{\sigma^{LR}} = j) = \begin{cases} \frac{1}{2} & \text{if } j = e - 1, \\ \frac{1}{2} & \text{if } j = e, \\ 0 & \text{otherwise.} \end{cases}$$

Thus according to Lemma 2,

$$\phi_i^C(\sigma_i^R, \sigma_i^{LR}, \sigma_i^R) = -\frac{\theta}{n} + \bar{d} + u(t_R - t_i) - u(t_{RE}^s - t_i),$$

while

$$\phi_i^C(\sigma_i^{LR}, \sigma_i^R, \sigma_i^{LR}) = \frac{\theta}{2n} + \frac{1}{2} \{u(|t_{LE}^s - t_i|) - u(t_{RE}^s - t_i)\},$$

and hence

$$\begin{aligned}
& \phi_i^C(\sigma_i^R, \sigma_i^{LR}, \sigma_i^R) - \phi_i^C(\sigma_i^{LR}, \sigma_i^R, \sigma_i^{LR}) \\
&= -\frac{3\theta}{2n} + \bar{d} + u(t_R - t_i) - \frac{1}{2} \{u(|t_{LE}^s - t_i|) + u(t_{RE}^s - t_i)\} \\
&= \frac{1}{2} \left[\bar{d} + u(t_R - t_i) - u(|t_{LE}^s - t_i|) - \frac{3\theta}{n} \right] \\
&+ \frac{1}{2} \left[\bar{d} + u(t_R - t_i) - u(t_{RE}^s - t_i) - \frac{3\theta}{n} \right] > 0
\end{aligned}$$

where each of the terms in brackets is strictly greater than 0 by Lemma 1, if $e \geq 2$ (and since $4e + 1 \leq n$ this implies $n \geq 9$).

Consider now the median voter, $i = m$.

One can calculate that

$$\begin{aligned}
\phi_m^C(\sigma_m^R, \sigma_m^{LR}, \sigma_{-m}^R) &= \frac{1}{2} \left\{ u(t_R - t_m) - u(t_{RE}^s - t_m) + \bar{d} - \frac{\theta}{n} \right\}, \\
\phi_m^C(\sigma_m^{LR}, \sigma_m^R, \sigma_{-m}^{LR}) &= 0.
\end{aligned}$$

Note that the second expression can also be obtained more directly by remembering that the median voter randomises between L and R if and only if

$$\mathbb{E} [U_i((L, \sigma_{-m}^{LR})) | (\emptyset, \emptyset)] = \mathbb{E} [U_i((R, \sigma_{-m}^{LR})) | (\emptyset, \emptyset)].$$

However, the expression for $\phi_m^C(\sigma_m^R, \sigma_m^{LR}, \sigma_{-m}^R)$ is strictly greater than 0, by Lemma 1.(iii) and the fact that $e \geq 1$ is always assumed.

This shows $\sigma^R \succ \sigma^{LR}$ in the presence of correlated errors. As already remarked, by Lemma 3, this also shows that σ^R dominates σ^{LR} according to maximum entropy in equilibria. The proof for $\sigma^L \succ \sigma^{LR}$ follows in exactly the same way and will lead to the same inequalities, *mutatis mutandis*, when considering $i \in \Omega^R \cup \{m\}$ this time.

(iv).

We prove that $\sigma^{LR} \succ \sigma^R$ according to the maximum entropy in strategies (version 1) criterion if n is large enough. That is, we need to show $\phi_i^C(\sigma_i^{LR}, \sigma_i^R, \frac{1}{2}\sigma_{-i}^{LR} + \frac{1}{2}\sigma_{-i}^R) \geq 0 \forall i \in \Omega$ (with a strict inequality for all i with $\sigma_i^{LR} \neq \sigma_i^R$ i.e. $i \in \Omega^L \cup \{m\}$). If $\sigma_i^{LR} = \sigma_i^R$, then $\phi_i^C(\sigma_i^{LR}, \sigma_i^R, \frac{1}{2}\sigma_{-i}^{LR} + \frac{1}{2}\sigma_{-i}^R) = 0$ so there is nothing to show.

Consider first $i \in \Omega^L$.

The distribution of the number of left-wing votes $\chi_{-i}^{\frac{1}{2}\sigma^{LR} + \frac{1}{2}\sigma^R}$ is given by:

$$p(\chi_{-i}^{\frac{1}{2}\sigma^{LR} + \frac{1}{2}\sigma^R} = j) = \begin{cases} \left(\frac{1}{2}\right)^{e-1} \left(\frac{3}{4}\right) & \text{if } j = 0, \\ \left(\frac{1}{2}\right)^{e-1} \left(\frac{1}{4}\binom{e-1}{j-1} + \frac{3}{4}\binom{e-1}{j}\right) & \text{if } 1 \leq j \leq e-1, \\ \left(\frac{1}{2}\right)^{e-1} \left(\frac{1}{4}\right) & \text{if } j = e, \\ 0 & \text{if } e+1 \leq j \leq 2e. \end{cases}$$

Then, by Lemma 2:

$$\begin{aligned} \phi_i^C \left(\sigma_i^{LR}, \sigma_i^R, \frac{1}{2}\sigma_{-i}^{LR} + \frac{1}{2}\sigma_{-i}^R \right) &= \phi_i^C \left(L, R, \frac{1}{2}\sigma_{-i}^{LR} + \frac{1}{2}\sigma_{-i}^R \right) \\ &= \frac{\theta}{n} \sum_{j=0}^{e-1} p(\chi_{-i}^{\frac{1}{2}\sigma^{LR} + \frac{1}{2}\sigma^R} = j) + p(\chi_{-i}^{\frac{1}{2}\sigma^{LR} + \frac{1}{2}\sigma^R} = e) [u(|t_{LE}^s - t_i|) - u(t_{RE}^s - t_i)] \\ &\quad + p(\chi_{-i}^{\frac{1}{2}\sigma^{LR} + \frac{1}{2}\sigma^R} = 0) [u(t_{RE}^s - t_i) - u(t_R - t_i) - \bar{d}] \\ &= \left(1 - \frac{1}{4} \left(\frac{1}{2}\right)^{e-1} \right) \frac{\theta}{n} + \frac{1}{4} \left(\frac{1}{2}\right)^{e-1} [u(|t_{LE}^s - t_i|) - u(t_{RE}^s - t_i)] \\ &\quad + \frac{3}{4} \left(\frac{1}{2}\right)^{e-1} [u(t_{RE}^s - t_i) - u(t_R - t_i) - \bar{d}]. \end{aligned}$$

Let

$$\begin{aligned} K_1 &= u(|t_{LE}^s - t_i|) - u(t_{RE}^s - t_i) \\ K_2 &= u(t_{RE}^s - t_i) - u(t_R - t_i) - \bar{d} \end{aligned}$$

Then since $i \in \Omega^L$, $K_1 \geq 0$, while by Lemma 1.(iii), $K_2 < 0$. Consider now the behaviour of $\phi_i^C \left(\sigma_i^{LR}, \sigma_i^R, \frac{1}{2}\sigma_{-i}^{LR} + \frac{1}{2}\sigma_{-i}^R \right)$ if we let $n, e \rightarrow \infty$ at an approximately constant ratio, that is, $e = \lfloor \frac{\tilde{s}_E n}{2} \rfloor$, for a fixed ‘‘ideal’’ share \tilde{s}_E . If we let $n \rightarrow \infty$ while e is fixed, the model is not very interesting in that then $s_E \rightarrow 0$ and eventually the coalition building assumptions made in the model, as summarised by Assumption 2, become very unrealistic. Under this latter asymptotic assumption, it is not possible to conclude anything about the sign of $\phi_i^C \left(\sigma_i^{LR}, \sigma_i^R, \frac{1}{2}\sigma_{-i}^{LR} + \frac{1}{2}\sigma_{-i}^R \right)$ as $n \rightarrow \infty$ for all $i \in \Omega^L$. This will depend on the exact values of the model-dependent parameters K_1 , K_2 , and θ . Thus, assume that $e = \lfloor \frac{\tilde{s}_E n}{2} \rfloor$. We have:

$$\lim_{n \rightarrow \infty} \phi_i^C \left(\sigma_i^{LR}, \sigma_i^R, \frac{1}{2}\sigma_{-i}^{LR} + \frac{1}{2}\sigma_{-i}^R \right) = 0.$$

This is intuitive. It reflects the fact that as the number of players increases, the expected difference in utility of a single vote goes to 0 as the probability of the vote

K_2	$N =$ minimal odd n	$e = \text{rd} \left(\frac{\tilde{s}_E N}{2} \right)$	$\phi(\cdot)$ at N
-0.5	45	6	1.03e-2
-1	53	7	3.98e-3
-3	69	9	5.69e-3
-10	85	11	4.44e-3

Table 3: Example calculations for $\phi_i^C \left(\sigma_i^{LR}, \sigma_i^R, \frac{1}{2}\sigma_{-i}^{LR} + \frac{1}{2}\sigma_{-i}^R \right)$ when $K_1 = K_1 \left(t_i = \frac{1}{2} \right) = 0$, K_2 is set to the value in the table, $\theta = 1$, and the ideal vote share $\tilde{s}_E = 0.25$. The N value in the table refers to the first odd integer where $\phi(\cdot) > 0$.

being pivotal tends to 0. However, as $\left(1 - \frac{1}{4} \left(\frac{1}{2} \right)^{e-1} \right) > 0 \quad \forall e \in \mathbb{N}$, and as the $\frac{\theta}{n}$ dominates the $2^{-e} = 2^{-\lfloor \frac{\tilde{s}_E n}{2} \rfloor}$ term in front of K_2 , for fixed values of K_1 , K_2 , and of the “ideal” parliament share \tilde{s}_E , we must have

$$\phi_i^C \left(\sigma_i^{LR}, \sigma_i^R, \frac{1}{2}\sigma_{-i}^{LR} + \frac{1}{2}\sigma_{-i}^R \right) > 0$$

for large enough values of n . Moreover, K_2 is decreasing in t_i , which implies that

$$K_2 \geq K_2 \left(t_i = \frac{1}{2} \right) = u \left(t_{RE}^s - \frac{1}{2} \right) - u \left(t_R - \frac{1}{2} \right) - \bar{d},$$

so that for fixed choices of t_R , t_{RE}^s , \bar{d} , $u(\cdot)$, and \tilde{s}_E , a large enough $N \in \mathbb{N}$ such that $\phi_i^C \left(\sigma_i^{LR}, \sigma_i^R, \frac{1}{2}\sigma_{-i}^{LR} + \frac{1}{2}\sigma_{-i}^R \right) > 0$ can be chosen uniformly over the player set Ω^L (recall $|\Omega^L| = e \rightarrow \infty$). Conversely however, given a number of players n , it is always possible to find values of K_2 such that $\phi_i^C \left(\sigma_i^{LR}, \sigma_i^R, \frac{1}{2}\sigma_{-i}^{LR} + \frac{1}{2}\sigma_{-i}^R \right) < 0$, that is the “large enough” n and e sizes must always depend on the underlying parameters of the model. In order to illustrate how large n and e must be, Tables 3 and 4 display the behaviour of $\phi_i^C \left(\sigma_i^{LR}, \sigma_i^R, \frac{1}{2}\sigma_{-i}^{LR} + \frac{1}{2}\sigma_{-i}^R \right)$ as a function of n for some possible parameter values. θ is set to 1, K_1 to $K_1 \left(t_i = \frac{1}{2} \right) = 0$, and \tilde{s}_E to either 0.25 or 0.15. e as function of n is rounded up or down, (rather than always down, as proposed in the text with $\lfloor \cdot \rfloor$). For all these parameters values, ϕ starts off negative for small n , crosses 0 and then decreases again, tending towards 0 as $n \rightarrow \infty$.

As can be seen, the required number of players N and e can become fairly large although when compared to the typical number of seats in a parliamentary assembly, they are well within normal bounds. On the other hand the values of ϕ are always quite small, so that the domination according to maximum entropy in strategies (version 1) is not very robust.

K_2	$N =$ minimal odd n	$e = \text{rd} \left(\frac{\tilde{s}_E N}{2} \right)$	$\phi(\cdot)$ at N
-0.5	75	6	1.51e-3
-1	101	8	4.02e-3
-3	127	10	3.48e-3
-10	155	12	2.79e-3

Table 4: Example calculations for $\phi_i^C \left(\sigma_i^{LR}, \sigma_i^R, \frac{1}{2}\sigma_{-i}^{LR} + \frac{1}{2}\sigma_{-i}^R \right)$ when $K_1 = K_1 \left(t_i = \frac{1}{2} \right) = 0$, K_2 is set to the value in the table, $\theta = 1$, and the ideal vote share $\tilde{s}_E = 0.15$. The N value in the table refers to the first odd integer where $\phi(\cdot) > 0$.

Consider now the median voter, $t_i = \frac{1}{2}$.

The distribution of the number of left-wing votes now changes. We have

$$\chi_{-m}^{\frac{1}{2}\sigma^{LR} + \frac{1}{2}\sigma^R} = B_{1/2}^e \sim \text{Bin} \left(e, \frac{1}{2} \right).$$

Also note that by the assumption that the different mixings of strategies are independent, we have

$$\begin{aligned} & \phi_m^C \left(\sigma_m^{LR}, \sigma_m^R, \frac{1}{2}\sigma_{-m}^{LR} + \frac{1}{2}\sigma_{-m}^R \right) \\ &= \frac{1}{2} \mathbb{E} \left[U_m \left(L, \frac{1}{2}\sigma_{-m}^{LR} + \frac{1}{2}\sigma_{-m}^R \right) \right] + \frac{1}{2} \mathbb{E} \left[U_m \left(R, \frac{1}{2}\sigma_{-m}^{LR} + \frac{1}{2}\sigma_{-m}^R \right) \right] \\ & \quad - \mathbb{E} \left[U_m \left(R, \frac{1}{2}\sigma_{-m}^{LR} + \frac{1}{2}\sigma_{-m}^R \right) \right] \\ &= \frac{1}{2} \phi_m^C \left(L, R, \frac{1}{2}\sigma_{-m}^{LR} + \frac{1}{2}\sigma_{-m}^R \right). \end{aligned}$$

Thus by Lemma 2:

$$\begin{aligned} & 2\phi_m^C \left(\sigma_m^{LR}, \sigma_m^R, \frac{1}{2}\sigma_{-m}^{LR} + \frac{1}{2}\sigma_{-m}^R \right) \\ &= \left(1 - \left(\frac{1}{2} \right)^e \right) \frac{\theta}{n} + \left(\frac{1}{2} \right)^e \left[u \left(t_{RE}^s - \frac{1}{2} \right) - u \left(t_R - \frac{1}{2} \right) - \bar{d} \right]. \end{aligned}$$

As for the $i \in \Omega^L$ case, we see that $K_2 = u \left(t_{RE}^s - \frac{1}{2} \right) - u \left(t_R - \frac{1}{2} \right) - \bar{d} < 0$, but because the exponential term tends to 0 much faster than the $\frac{\theta}{n}$ term, for large enough values of n and e , the above will be strictly greater than 0. This shows $\sigma^{LR} \succ \sigma^R$ according to the maximum entropy in strategies (version 1) criterion.

The proof of $\sigma^{LR} \succ \sigma^L$ is analagous, *mutatis mutandis*. That is, for $i \in \Omega^R$, we have

$$\begin{aligned} & \phi_i^C \left(\sigma_i^{LR}, \sigma_i^L, \frac{1}{2}\sigma_{-i}^{LR} + \frac{1}{2}\sigma_{-i}^L \right) \\ &= \left(1 - \frac{1}{4} \left(\frac{1}{2} \right)^{e-1} \right) \frac{\theta}{n} - \frac{1}{4} \left(\frac{1}{2} \right)^{e-1} [u(t_i - t_{LE}^s) - u(|t_{RE}^s - t_i|)] \\ &\quad - \frac{3}{4} \left(\frac{1}{2} \right)^{e-1} [u(t_i - t_L) - u(|t_{LE}^s - t_i|) + \bar{d}], \end{aligned}$$

while for $i = m$, $\chi_{-m}^{\frac{1}{2}\sigma^{LR} + \frac{1}{2}\sigma^L} = e + B_{1/2}^e$, so that

$$\begin{aligned} & 2\phi_m^C \left(\sigma_m^{LR}, \sigma_m^L, \frac{1}{2}\sigma_{-m}^{LR} + \frac{1}{2}\sigma_{-m}^L \right) \\ &= \mathbb{E} \left[U_m \left(R, \frac{1}{2}\sigma_{-m}^{LR} + \frac{1}{2}\sigma_{-m}^L \right) \right] - \mathbb{E} \left[U_m \left(L, \frac{1}{2}\sigma_{-m}^{LR} + \frac{1}{2}\sigma_{-m}^R \right) \right] \\ &= \left(1 - \left(\frac{1}{2} \right)^e \right) \frac{\theta}{n} - \left(\frac{1}{2} \right)^e \left[u \left(\frac{1}{2} - t_L \right) - u \left(\frac{1}{2} - t_{LE}^s \right) + \bar{d} \right]. \end{aligned}$$

(ii).

We show that for every set of model parameters, “ideal parliamentary seat-share” \tilde{s}_E and $\bar{t} \in (\frac{1}{2}, 1]$, there exists an $N \in \mathbb{N}$ large enough, such that $\forall n > N$, $e \geq e(n) = \lfloor \frac{\tilde{s}_E n}{2} \rfloor$, and $i \in \Omega^L \cup \{i \in \Omega^R : t_i \geq \bar{t}\}$, the function

$$f_i(\varepsilon) = \phi_i^C (\sigma_i^{LR}, \sigma_i^L, (1 - \varepsilon)\sigma_{-i}^{LR} + \varepsilon\sigma_{-i}^L)$$

is monotonically decreasing in ε . In particular, it satisfies the near-monotonicity property defined in Lemma 4.(ii), that is, that $f_i(\varepsilon) \geq f_i(\frac{1}{2}) \quad \forall \varepsilon \in [0, \frac{1}{2}]$, and $f_i(\varepsilon) \leq f_i(\frac{1}{2}) \quad \forall \varepsilon \in [\frac{1}{2}, 1]$, and hence $\sigma^{LR} \succsim_i^\perp \sigma^L$ is equivalent to σ^{LR} dominating σ^L according to maximum entropy in strategies (version 1) for these players (and both dominations are strict in the case of the set of players $\{i \in \Omega^R : t_i \geq \bar{t}\}$).

Firstly, if $i \in \Omega^L$, $\sigma_i^L = \sigma_i^{LR}$, so

$$\phi_i^C (\sigma_i^{LR}, \sigma_i^L, (1 - \varepsilon)\sigma_{-i}^{LR} + \varepsilon\sigma_{-i}^L) = 0 \quad \forall \varepsilon \in [0, 1],$$

and there is nothing more to show.

Thus consider $i \in \Omega^R$.

When all players except i play $(1-\varepsilon)\sigma_{-i}^{LR} + \varepsilon\sigma_{-i}^L$, the median voter has a probability $\frac{1-\varepsilon}{2}$ of voting R and a probability $\frac{1+\varepsilon}{2}$ of voting L , while the $e-1$ remaining R conventional voters' L votes are distributed according to B_ε^{e-1} . Hence,

$$p\left(\chi_{-i}^{(1-\varepsilon)\sigma_{-i}^{LR} + \varepsilon\sigma_{-i}^L} = j\right) = \begin{cases} 0 & \text{if } j < e, \\ \frac{1-\varepsilon}{2}p(B_\varepsilon^{e-1} = 0) & \text{if } j = e, \\ \frac{1+\varepsilon}{2}p(B_\varepsilon^{e-1} = j - e - 1) \\ \quad + \frac{1-\varepsilon}{2}p(B_\varepsilon^{e-1} = j - e) & \text{if } e + 1 \leq j \leq 2e - 1, \\ \frac{1+\varepsilon}{2}p(B_\varepsilon^{e-1} = e - 1) & \text{if } j = 2e. \end{cases}$$

Using Lemma 2, we calculate:

$$\begin{aligned} \phi_i^C(\sigma_i^{LR}, \sigma_i^L, (1-\varepsilon)\sigma_{-i}^{LR} + \varepsilon\sigma_{-i}^L) &= -\phi_i^C(L, R, (1-\varepsilon)\sigma_{-i}^{LR} + \varepsilon\sigma_{-i}^L) \\ &= \frac{\theta}{n} \left(1 - \frac{(1-\varepsilon)^e}{2}\right) - \frac{(1-\varepsilon)^e}{2} [u(t_i - t_{LE}^s) - u(|t_i - t_{RE}^s|)] \\ &\quad - \left(\frac{1+\varepsilon}{2}\right) \varepsilon^{e-1} [u(t_i - t_L) - u(t_i - t_{LE}^s) + \bar{d}] \\ &= \frac{\theta}{n} \left(1 - \frac{(1-\varepsilon)^e}{2}\right) + (1-\varepsilon)^e \frac{K_1}{2} + (1+\varepsilon)\varepsilon^{e-1} \frac{K_2}{2}, \end{aligned}$$

where this time we define

$$\begin{aligned} K_1 &:= u(|t_i - t_{RE}^s|) - u(t_i - t_{LE}^s), \\ K_2 &:= u(t_i - t_{LE}^s) - u(t_i - t_L) - \bar{d}. \end{aligned}$$

Since $i \in \Omega^R$, we have $K_1 > 0$ and $K_2 < 0$ (by the Polarization Corollary). This means that the first $\frac{\theta}{n}$ term is increasing in ε , while the two other terms are decreasing. Differentiating w.r.t. ε yields:

$$\begin{aligned} \frac{\partial}{\partial \varepsilon} \phi_i^C(\sigma_i^{LR}, \sigma_i^L, (1-\varepsilon)\sigma_{-i}^{LR} + \varepsilon\sigma_{-i}^L) &= \frac{\theta}{n} \frac{e}{2} (1-\varepsilon)^{e-1} - \frac{eK_1}{2} (1-\varepsilon)^{e-1} + \frac{(2e-1)K_2}{2} (1+\varepsilon)\varepsilon^{e-2} \\ &= \frac{e}{2} (1-\varepsilon)^{e-1} \left(\frac{\theta}{n} - K_1\right) + \frac{(2e-1)K_2}{2} (1+\varepsilon)\varepsilon^{e-2} \end{aligned}$$

Suppose that $K_1 > 0$ is fixed. Then for large enough n , the above derivative will be negative, and $\phi_i^C(\sigma_i^{LR}, \sigma_i^L, (1-\varepsilon)\sigma_{-i}^{LR} + \varepsilon\sigma_{-i}^L)$ must be monotonically decreasing. The issue is that as the number of players increases, depending on the asymptotic distributional assumptions of t_i , there may be players which have their ideal point arbitrarily close to $\frac{1}{2}$ so that for those players $K_1 \rightarrow 0$. In other words, unlike the

case above where $\varepsilon = \frac{1}{2}$, it is not possible to find an N large enough, such that the derivative

$$\frac{\partial}{\partial \varepsilon} \phi_i^C (\sigma_i^{LR}, \sigma_i^L, (1 - \varepsilon)\sigma_{-i}^{LR} + \varepsilon\sigma_{-i}^L) < 0 \quad \forall n \geq N \quad \forall i \in \Omega^R(n).$$

In order to be able to make statements about the derivative above, and hence the monotonicity of $f_i(\varepsilon) = \phi_i^C (\sigma_i^{LR}, \sigma_i^L, (1 - \varepsilon)\sigma_{-i}^{LR} + \varepsilon\sigma_{-i}^L)$ w.r.t. ε , uniformly for a subset of players in Ω^R , we must restrict the set of player ideal policies to a compact subset. Suppose that we only consider $t_i \in [\bar{t}, 1]$. In that case $K_1(t_i) \geq K_1(\bar{t}) \quad \forall t_i \in [\bar{t}, 1]$ and hence if n and e are large enough, for all those players $f_i(\varepsilon)$ will be uniformly decreasing. This proves the second part of the assertion in (ii).

Suppose now that we may not make compactness assumptions about the set $\{t_i : i \in \Omega^R\}$. We show that $\phi_i^C (\sigma_i^{LR}, \sigma_i^L, (1 - \varepsilon)\sigma_{-i}^{LR} + \varepsilon\sigma_{-i}^L) \geq 0 \quad \forall \varepsilon \in [0, \frac{1}{2}]$, that is, that σ^{LR} is immune to uncorrelated errors at level $\frac{1}{2}$ w.r.t. σ^L for any player $i \in \Omega^R$. By Lemma 4.(i), this implies that $\sigma^{LR} \succ_i^{\frac{1}{2}} \sigma^L \quad \forall i \in \Omega^R$. Consider

$$\begin{aligned} & \frac{\partial^2}{\partial \varepsilon^2} \phi_i^C (\sigma_i^{LR}, \sigma_i^L, (1 - \varepsilon)\sigma_{-i}^{LR} + \varepsilon\sigma_{-i}^L) \\ &= \frac{-(e-1)e}{2} (1 - \varepsilon)^{e-2} \left(\frac{\theta}{n} - K_1 \right) + \frac{(2e-1)(2e-3)K_2}{2} (1 + \varepsilon)\varepsilon^{e-3} \end{aligned}$$

Now suppose that $(\frac{\theta}{n} - K_1) \leq 0$. Then the argument made above is applicable, and

$$\frac{\partial}{\partial \varepsilon} \phi_i^C (\sigma_i^{LR}, \sigma_i^L, (1 - \varepsilon)\sigma_{-i}^{LR} + \varepsilon\sigma_{-i}^L) < 0 \quad \forall \varepsilon \in [0, 1]$$

so that $f_i(\varepsilon)$ is monotonically decreasing. Suppose that $(\frac{\theta}{n} - K_1) \geq 0$. Then

$$\frac{\partial^2}{\partial \varepsilon^2} \phi_i^C (\sigma_i^{LR}, \sigma_i^L, (1 - \varepsilon)\sigma_{-i}^{LR} + \varepsilon\sigma_{-i}^L) < 0 \quad \forall \varepsilon \in [0, 1]$$

This implies that the derivative $\frac{\partial}{\partial \varepsilon} f_i(\varepsilon)$ can have at most one root. Now $f_i(\varepsilon = 0) = \frac{\theta}{2n} + \frac{K_1}{2} > 0$. Thus even if $f_i(\varepsilon)$ is first increasing, it can have at most one root itself. But by (iv) above, $f_i(\varepsilon = \frac{1}{2}) > 0$ if n and e are large enough. Thus, the zero of $f_i(\varepsilon)$ must occur after $\frac{1}{2}$. This shows the immunity at level $\varepsilon^* = \frac{1}{2} \quad \forall i \in \Omega^R$.

Now consider the median voter $i = m$.

As above, we show that

$$f_m(\varepsilon) = \phi_m^C (\sigma_m^{LR}, \sigma_m^L, (1 - \varepsilon)\sigma_{-m}^{LR} + \varepsilon\sigma_{-m}^L) \geq 0 \quad \forall \varepsilon \in \left[0, \frac{1}{2} \right]$$

We have $\chi_{-m}^{(1-\varepsilon)\sigma^{LR} + \varepsilon\sigma^L} = e + B_\varepsilon^e$. By Lemma 2, we calculate:

$$\begin{aligned} 2\phi_m^C(\sigma_m^{LR}, \sigma_m^L, (1-\varepsilon)\sigma_{-m}^{LR} + \varepsilon\sigma_{-m}^L) &= -\phi_i^C(L, R, (1-\varepsilon)\sigma_{-m}^{LR} + \varepsilon\sigma_{-m}^L) \\ &= (1 - (1-\varepsilon)^e) \frac{\theta}{n} + \varepsilon^e \left[u\left(\frac{1}{2} - t_{LE}^s\right) - u\left(\frac{1}{2} - t_L\right) - \bar{d} \right] \\ &= (1 - (1-\varepsilon)^e) \frac{\theta}{n} + \varepsilon^e K. \end{aligned}$$

where $K := u\left(\frac{1}{2} - t_{LE}^s\right) - u\left(\frac{1}{2} - t_L\right) - \bar{d} < 0$, by the Polarization Corollary. We have:

$$\begin{aligned} \frac{\partial}{\partial \varepsilon} f_m(\varepsilon) &= e \left\{ \frac{\theta}{n} (1-\varepsilon)^{e-1} + \varepsilon^{e-1} K \right\} < 0 \\ \Leftrightarrow \frac{\theta}{nK} &< \left(\frac{\varepsilon}{1-\varepsilon} \right)^{e-1}, \end{aligned}$$

which for large enough n and e is true if and only if $\varepsilon \geq \frac{1}{2}$, since $\frac{\theta}{nK}$ dominates the exponential term in e . However, note that $f_m(\varepsilon = 0) = 0$. In particular, this shows $f_m(\varepsilon) \geq 0 \forall \varepsilon \in [0, \frac{1}{2}]$ and by Lemma 4.(i) concludes the proof of $\sigma^{LR} \succ^\perp \sigma^L$.

The proof for the $\sigma^{LR} \succ \sigma^R$ case follows either by symmetry (recall that $C = (\emptyset, \emptyset)$ so the set of possibilities and players is perfectly symmetric) or by the same arguments we made above, but applied to the function $\phi_i^C(\sigma_i^{LR}, \sigma_i^R, (1-\varepsilon)\sigma_{-i}^{LR} + \varepsilon\sigma_{-i}^R)$.

A.2 Proof of Proposition 7

(i) and (iii).

By Lemma 3, 'domination in the presence of correlated errors' and 'domination according to maximum entropy in equilibria' are both equivalent to

$$\phi_i^C(\sigma_i^R, \sigma_i^L, \sigma_{-i}^R) - \phi_i^C(\sigma_i^L, \sigma_i^R, \sigma_{-i}^L) > 0 \quad \forall i \in \Omega,$$

since in this case, $\sigma_i^R \neq \sigma_i^L$ for every $i \in \Omega$. We calculate (directly or using Lemma 2):

$$\begin{aligned} \mathbb{E}[U_i(R, \sigma_{-i}^R)] &= u(|t_R - t_i|) - \theta \left(\frac{1}{2} + \frac{1}{2n} \right), \\ \mathbb{E}[U_i(L, \sigma_{-i}^R)] &= u(|t_{RE}^w - t_i|) - \theta \left(\frac{1}{2} - \frac{1}{2n} \right) - \bar{d}, \\ \mathbb{E}[U_i(L, \sigma_{-i}^L)] &= u(|t_L - t_i|) - \theta \left(\frac{1}{2} + \frac{1}{2n} \right), \\ \mathbb{E}[U_i(R, \sigma_{-i}^L)] &= u(|t_{RE}^w - t_i|) - \theta \left(\frac{1}{2} - s_E + \frac{1}{2n} \right) - \bar{d}, \end{aligned}$$

and

$$\phi_i^C(\sigma_i^R, \sigma_i^L, \sigma_{-i}^R) - \phi_i^C(\sigma_i^L, \sigma_i^R, \sigma_{-i}^L) = u(|t_R - t_i|) - u(|t_L - t_i|) - \frac{\theta}{n} + s_E \theta.$$

We see that if $i \in \Omega^R \cup \{m\}$ and n is large enough, this is always greater than 0, and so $\sigma^R \succ_i^{\parallel} \sigma^L$. If $i \in \Omega^L$ however, $u(|t_R - t_i|) - u(|t_L - t_i|) < 0$, and so we need the PCC introduced in this section. In that case, for large enough n , we also obtain $\sigma^R \succ_i^{\parallel} \sigma^L$.

(iv).

We need to show that

$$\phi_i^C \left(\sigma_i^L, \sigma_i^R, \frac{1}{2}\sigma_{-i}^L + \frac{1}{2}\sigma_{-i}^R \right) > 0 \quad \forall i \in \Omega.$$

For any $i \in \Omega$, $\chi_{-i}^{\frac{1}{2}\sigma^L + \frac{1}{2}\sigma^R} = B_{1/2}^{2e}$. Thus by Lemma 2,

$$\phi_i^C \left(\sigma_i^L, \sigma_i^R, \frac{1}{2}\sigma_{-i}^L + \frac{1}{2}\sigma_{-i}^R \right) = \left(\frac{1}{2} \right)^{2e} \left[u(|t_L - t_i|) - u(|t_R - t_i|) - s_E \theta - \frac{\theta}{n} \right] + \frac{\theta}{n}$$

We argue as in the proof of Proposition 5.(iv). If $e \geq e(n) := \lfloor \frac{\tilde{s}_E n}{2} \rfloor$, then the $(\frac{1}{2})^{2e}$ is dominated by $\frac{\theta}{n}$ for large enough n , so that indeed $\phi_i^C(\sigma_i^L, \sigma_i^R, \frac{1}{2}\sigma_{-i}^L + \frac{1}{2}\sigma_{-i}^R) > 0 \forall i \in \Omega$, since uniformly in i :

$$u(|t_L - t_i|) - u(|t_R - t_i|) - s_E \theta - \frac{\theta}{n} \geq u(|t_L - t_R^\Omega|) - u(|t_R - t_R^\Omega|) - s_E \theta - \frac{\theta}{n}.$$

(ii).

We proceed to calculate $\phi_i^C(\sigma_i^L, \sigma_i^R, (1 - \varepsilon)\sigma_{-i}^L + \varepsilon\sigma_{-i}^R)$. We have $\chi_{-i}^{(1-\varepsilon)\sigma^L + \varepsilon\sigma^R} = B_{1-\varepsilon}^{2e}$.

Thus, by Lemma 2:

$$\begin{aligned} & \phi_i^C(\sigma_i^L, \sigma_i^R, (1 - \varepsilon)\sigma_{-i}^L + \varepsilon\sigma_{-i}^R) \\ &= \varepsilon^{2e} [u(|t_{RE}^w - t_i|) - u(|t_R - t_i|) - \bar{d}] \\ & \quad + (1 - \varepsilon)^{2e} \left[u(|t_L - t_i|) - u(|t_{RE}^w - t_i|) + \bar{d} - \left(s_E + \frac{1}{n} \right) \theta \right] + \frac{\theta}{n} \\ &= \varepsilon^{2e} K_1(t_i) + (1 - \varepsilon)^{2e} K_2(t_i) + \frac{\theta}{n}, \end{aligned}$$

where we let

$$\begin{aligned} K_1(t_i) &:= u(|t_{RE}^w - t_i|) - u(|t_R - t_i|) - \bar{d} \\ K_2(t_i) &:= u(|t_L - t_i|) - u(|t_{RE}^w - t_i|) + \bar{d} - \left(s_E + \frac{1}{n} \right) \theta \end{aligned}$$

By the Polarization Corollary, $K_1(t_i) < 0 \forall i \in \Omega$, while $K_2(t_i) \geq -\frac{\theta}{n}$. Moreover, as in the proof of the Polarization Lemma (Lemma 1), one sees that

$$\begin{aligned}\frac{\partial}{\partial t_i} K_2(t_i) &= u'(t_i - t_L) - u'(t_i - t_{RE}^w) \leq 0 & \text{if } t_i \in (t_{RE}^w, 1] \\ \frac{\partial}{\partial t_i} K_2(t_i) &= u'(t_i - t_L) + u'(t_{RE}^w - t_i) < 0 & \text{if } t_i \in (t_L, t_{RE}^w) \\ \frac{\partial}{\partial t_i} K_2(t_i) &= u'(t_{RE}^w - t_i) - u'(t_L - t_i) \leq 0 & \text{if } t_i \in [0, t_L)\end{aligned}$$

In particular, by the mean-value theorem, this implies that if $t_i < t_{RE}^w$, then from the Weak PC we get:

$$\bar{d} > s_E \theta + u(|t_{RE}^w - t_i|) - u(|t_L - t_i|),$$

and $K_2(t_i) > 0$ if n is large enough. In turn, it follows from this that if $t_i < t_{RE}^w$ then $f_i(\varepsilon) := \phi_i^C(\sigma_i^L, \sigma_i^R, (1 - \varepsilon)\sigma_{-i}^L + \varepsilon\sigma_{-i}^R)$ is monotonically decreasing in ε , and by Lemma 4.(ii), 'domination according to maximum entropy in strategies' is equivalent to 'domination in the presence of uncorrelated errors'.

On the other hand, if $t_i \geq t_{RE}^w$, it is possible that $K_2(t_i) < 0 \forall n \in \mathbb{N}$. In that case, we must argue manually that $f_i(\varepsilon) > 0 \forall \varepsilon \in [0, \frac{1}{2}]$ for large enough n . Indeed, we always have:

$$\begin{aligned}\frac{\partial}{\partial t_i} f_i(\varepsilon) &= 2e \{ \varepsilon^{2e-1} K_1(t_i) - (1 - \varepsilon)^{2e-1} K_2(t_i) \} \\ \frac{\partial^2}{\partial t_i^2} f_i(\varepsilon) &= 2e(2e - 1) \{ \varepsilon^{2e-2} K_1(t_i) + (1 - \varepsilon)^{2e-2} K_2(t_i) \}.\end{aligned}$$

Analogously to the proof of Proposition 5 (ii), we see that if $K_2(t_i) \geq 0$, then as in the $t_i < t_{RE}^w$ case, $f_i(\varepsilon)$ is monotonically decreasing. If $K_2(t_i) < 0$, $\frac{\partial^2}{\partial t_i^2} f_i(\varepsilon) < 0 \forall \varepsilon \in [0, 1]$, both $\frac{\partial}{\partial t_i} f_i(\varepsilon)$ and $f_i(\varepsilon)$ may have only one root in $[0, 1]$. But together with

$$\begin{aligned}f_i(\varepsilon = 0) &= K_2(t_i) + \frac{\theta}{n} \geq 0, \\ f_i\left(\varepsilon = \frac{1}{2}\right) &> 0 \text{ if } n \text{ large enough, by (iv) above,}\end{aligned}$$

this implies $f_i(\varepsilon) > 0 \forall \varepsilon \in [0, \frac{1}{2}]$. By Lemma 4 (i) this proves that if n and e are large enough, $\sigma^L \succ_i^{\frac{1}{2}} \sigma^R$ for all $i \in \Omega$.

Online Appendix

B Real World Examples of CPPs and of Extremist Parties in Government

Typically two large center-right and center-left parties vied for majorities in parliament. In some cases, when they did not have enough seats to govern on their own, these conventional parties had to rely on smaller, often single-issue parties, and competed with each other to form awkward coalition or minority governments.³³ This has become more acute over the last decade since the financial crisis, as the traditional postwar order of two centrist conventional blocks dominating politics has collapsed. Parties with more extreme policy agendas made especially large gains in connection with the recent migrant crisis, when more than a million refugees entered the EU. Conventional parties sometimes try to distinguish themselves from the other moderate party by promising not to go into coalition with extreme parties. However such promises are often broken.³⁴ Even when no explicit promises are made before the election, small extreme parties are able to extract strong concessions from conventional parties and are able to shape policy to a far greater extent than their vote share would justify, potentially hurting the utilitarian welfare of the electorate.

One example is the UK, where after the 2017 general election, neither Labour and the Scottish National Party (262 and 35 out of 650 seats respectively), nor the Conservatives (317 seats) could form a government. The Conservatives chose to rely on the support of the Northern Irish Democratic Unionist Party's (DUP) 10 MPs in order to obtain a majority (Apostolova et al., 2019). The DUP promotes some fairly ex-

³³In the 2024 elections in the German states of Thuringia and Saxony, the center-right party CDU and the center-left party SPD precluded via CPPs before the election to form a coalition government with the extremist rightwing party AfD. In order to keep this promise, the two center parties included the populist party BSW (Buendnis Sarah Wagenknecht) in their government coalition in Thuringia and formed a minority government in Saxony. Another example is Spain, where since the re-instatement of democracy, the People's Party (PP) and the Socialist Party (PSOE) have had to rely on Basque and Catalan nationalist parties, conceding much more than the vote-share of these nationalist parties would imply. Or consider Israel, where the Likud party has been heavily dependent on ultra-orthodox parties for many years.

³⁴Recently in Estonia's 2019 national parliamentary election, for example, where the Centre Party broke its pre-election promise and entered into coalition with the far-right Conservative People's Party (EKRE).

treme cultural policies, when placed within the UK political spectrum,³⁵ such as not recognizing same-sex marriage and considerably reducing abortion rights.³⁶ One of its representatives in the Northern Ireland assembly is openly creationist and supports creationist teaching in school (Ainsworth, 2016; BBC News, 2018b). It has also been a factor in preventing a resolution to Brexit as it categorically rejects any possibility of custom checks between Northern Ireland and the rest of the UK.

Another example is New Zealand. In the 2017 general election for instance, the centre-right National Party received 56 of the 120 seats in New Zealand’s parliament, the centre-left Labour 46, the Greens 8, the extreme nationalist party New Zealand First 9, and the pro-business (classically liberal) centre-right ACT party 1 (Electoral Commission, 2017). Labour and the Greens, and the National Party and ACT respectively, were widely expected to work together in government formation. New Zealand First was thus in the position of sole kingmaker, with several fairly extreme demands for any coalition that wished to have their support, such as reducing net immigration to 10’000 people or less, two national referenda on the removal of the two special indigenous Maori’s seats and reducing the size of parliament to 100 (now party website mentions aim to reduce parliament to 80), and moving Auckland’s main commercial port further from the city (Burr, 2017; New Zealand First, 2019). On October 19th 2017, Labour, the Greens and New Zealand First announced that they had come to a coalition agreement (Barraclough, 2017).

Other examples include Catalonia, recently Austria in 2017, where the FPÖ and ÖVP have formed a governing coalition, as well as the traditional dilemma of the Greens and Social-Democrats (SPD) in Germany, when they would have a majority by entering into coalition with Die Linke.

C Further Proofs

Before providing the proofs of the Propositions in the main text, we introduce several helpful abbreviations and smaller results that will be used throughout the proofs that

³⁵As the DUP is currently the strongest party in the local government councils of Northern Ireland and tied with Sinn Féin in the Northern Ireland Assembly, its positions can hardly be described as extreme within the Northern Irish political spectrum.

³⁶Interestingly, these positions are not present in any of the DUP’s recent manifestos. However, see BBC News (2015) for the most recent vote on legalizing same-sex marriage in Northern Ireland, and Birchard (2000) and BBC News (2018a) regarding abortion.

follow.

We start by introducing $\eta = t_R^\Omega - t_R = t_L - t_L^\Omega$, which relates the preferred policies in the \mathcal{T} -policy dimension of the conventional voters who are most different from the median voter to the preferences of the conventional parties. Note that η can be either positive or negative.

Second, we define the party polarization of the voting game as the difference between the preferred platforms of the two parties, i.e.,

$$\pi := t_R - t_L. \quad (11)$$

Note that $\eta \geq -\frac{\pi}{2}$. Second, we define the (lack of) strong bargaining power of conventional parties as the difference between what they can obtain in the \mathcal{T} -dimension when bargaining with E in the case where the latter has strong bargaining power w.r.t. to their platforms, i.e.,

$$\beta^s := t_R - t_{RE}^s = t_{LE}^s - t_L. \quad (12)$$

When β^s is large, the conventional parties are not able to shift the policy in the \mathcal{T} -dimension towards their preferred policy when bargaining with an extreme party that has strong bargaining power. Third, we similarly define the (lack of) weak bargaining power of conventional parties as the difference between what they can obtain in the \mathcal{T} -dimension when bargaining with E in the case where the latter has weak bargaining power w.r.t. to their platforms, i.e.,

$$\beta^w := t_R - t_{RE}^w = t_{LE}^w - t_L. \quad (13)$$

When β^w is large, the conventional parties are not able to shift the policy in the \mathcal{T} -dimension towards their preferred policy when bargaining with an extreme party that has weak bargaining power. Fourth, we introduce the parameter γ to capture the share of votes that E requires in order to obtain a majority.

$$\gamma := \frac{1}{2} - s_E. \quad (14)$$

Throughout the paper, we assume that the disutility from adopting the extreme policy $d = \bar{d}$ is sufficiently large compared to policy differences in the left-right dimension and perks. This is captured in Assumption 1(ii) and with the parameters introduced above it can be rewritten as:

Polarization Conditions (PCs)

$$(Weak) : \quad \bar{d} \geq s_E \theta + u(|\beta^w + \eta|) - u(\pi + \eta).$$

$$(Strong) : \quad \bar{d} \geq s_E \theta + u(|\beta^s + \eta|) - u(\pi + \eta).$$

Recall that the conditions (PCs) imply that any strategically-voting conventional voter would rather see a single conventional party government, even if formed by the other party which he does not support, than a coalition including E , which would implement \bar{d} .³⁷ In other words, an intuitive formulation of the polarization conditions is that conventional voters care more about policy differences in the extreme dimension than about differences in the conventional left-right dimension. Indeed, we have the following result.

Lemma 1 (Polarization)

Let $j \in \{w, s\}$ refer to the level of bargaining power of the extreme party. Then, the PCs imply that for all $i \in \Omega$:

$$(i) \quad \bar{d} \geq s_E \theta \pm \left(u(|t_{LE}^j - t_i|) - u(|t_R - t_i|) \right).$$

$$(ii) \quad \bar{d} \geq s_E \theta \pm \left(u(|t_{LE}^j - t_i|) - u(|t_L - t_i|) \right).$$

$$(iii) \quad \bar{d} \geq s_E \theta \pm \left(u(|t_{RE}^j - t_i|) - u(|t_R - t_i|) \right).$$

$$(iv) \quad \bar{d} \geq s_E \theta \pm \left(u(|t_{RE}^j - t_i|) - u(|t_L - t_i|) \right).$$

If $t_{LE}^j \neq t_{RE}^j$ and hence $t_L \neq t_R$ then (ii) and (iii) hold strictly for all $i \in \Omega$.

Moreover, we consider the following upper-bound for the sensitivity of voters to perks:

Sensitivity-to-Perks Condition (SPC)

$$\begin{aligned} \theta \left(\gamma - \frac{1}{2n} \right) &< u \left(t_R^\Omega - t \left(\frac{1}{2} - \frac{1}{2n}, \gamma + \frac{1}{2n} \right) \right) - u(\pi + \eta) \\ &= u \left(t \left(\gamma + \frac{1}{2n}, \frac{1}{2} - \frac{1}{2n} \right) - t_L^\Omega \right) - u(\pi + \eta), \end{aligned}$$

³⁷Note that if we assume that $\beta^w + \eta > 0$, that is, that $t_L^\Omega < t_{LE}^w$ and conversely $t_{RE}^w < t_R^\Omega$, then the Weak Polarization Condition implies the Strong Condition since u is assumed to be decreasing.

where $t(s_L, s_R)$ is the Nash bargaining solution defined in Assumption 2.(iii). Note that the SPC depends on n , both via the Nash bargaining solution and directly on the left-hand side. However, we have:

$$\begin{aligned} \theta \left(\gamma - \frac{1}{2n} \right) &< \theta \gamma \quad \text{and} \\ u \left(t_R^\Omega - t \left(\frac{1}{2}, \gamma \right) \right) &< u \left(t_R^\Omega - t \left(\frac{1}{2} - \frac{1}{2n}, \gamma + \frac{1}{2n} \right) \right), \end{aligned}$$

since increasing the share of parliamentary seats accruing to L will bring $t(s_L, s_R)$ closer to t_L , that is $t(s_L, s_R)$ is decreasing in s_L . In turn, the above inequalities imply that the SPC is implied by the following stronger condition:

$$\theta \gamma < u \left(t_R^\Omega - t \left(\frac{1}{2}, \gamma \right) \right) - u(\pi + \eta). \quad (15)$$

We keep the n -dependent SPC as the main condition in the paper however, as the proof of the coming Proposition 1 in the Appendix shows that the n -dependent condition is both sufficient and necessary for σ^L and σ^R to not be equilibria of our stage 2 voting subgame if $C = (E, E)$. Were we to use the (only sufficient) inequality in (15), it would be possible that even though (15) does *not* hold, σ^L and σ^R are still not equilibria of the subgame.³⁸

An intuitive formulation of both conditions is that voters care less about perks than about policy differences in the conventional left-right dimension.

We are now in a position to find the equilibria of the voting subgame.

C.1 Details on Equilibrium Selection

Recall that

$$t(s_L, s_R) = \operatorname{argmax}_{t \in [t_L, t_R]} s_L u(t - t_L) + s_R u(t_R - t)$$

is the policy which is implemented when both conventional parties negotiate about their grand-coalition platform. Then the differences in expected utilities for voting for L versus R given the voting strategies of the other voters as defined by $\phi_i^C(\sigma'_i, \sigma''_i, \sigma_{-i})$ in (8) takes the following form.

³⁸That is, while $\theta \left(\gamma - \frac{1}{2n} \right) \geq u \left(t_R^\Omega - t \left(\frac{1}{2} - \frac{1}{2n}, \gamma + \frac{1}{2n} \right) \right) - u(\pi + \eta) \Rightarrow \sigma^L, \sigma^R$ are equilibria, $\theta \gamma \geq u \left(t_R^\Omega - t \left(\frac{1}{2}, \gamma \right) \right) - u(\pi + \eta)$ carries no such implication. When n is large enough however, it is clear that the strict inequality $\theta \gamma > u \left(t_R^\Omega - t \left(\frac{1}{2}, \gamma \right) \right) - u(\pi + \eta)$ would be sufficient for σ^L and σ^R to be equilibria if $C = (E, E)$.

Lemma 2

The following assertions hold for every $i \in \Omega$.³⁹

(a) If $C = (\emptyset, \emptyset)$, then

$$\begin{aligned} \phi_i^C(L, R, \sigma_{-i}) &= \frac{1}{n} \theta \left[\sum_{j=0}^{e-1} p(\mathcal{X}_{-i}^\sigma = j) - \sum_{j=e+1}^{2e} p(\mathcal{X}_{-i}^\sigma = j) \right] \\ &\quad + p(\mathcal{X}_{-i}^\sigma = e) [u(|t_{LE}^s - t_i|) - u(|t_{RE}^s - t_i|)] \\ &\quad + p(\mathcal{X}_{-i}^\sigma = 0) [u(|t_{RE}^s - t_i|) - u(|t_R - t_i|) - \bar{d}] \\ &\quad + p(\mathcal{X}_{-i}^\sigma = 2e) [u(|t_L - t_i|) - u(|t_{LE}^s - t_i|) + \bar{d}]. \end{aligned}$$

(b) If $C = (E, \emptyset)$, then

$$\begin{aligned} \phi_i^C(L, R, \sigma_{-i}) &= p(\mathcal{X}_{-i}^\sigma = 0) [u(|t_{RE}^w - t_i|) - u(|t_R - t_i|) - \bar{d}] \\ &\quad + p(\mathcal{X}_{-i}^\sigma = 2e) \left[u(|t_L - t_i|) - u(|t_{RE}^w - t_i|) + \bar{d} - \left(s_E + \frac{1}{n} \right) \theta \right] \\ &\quad + \frac{1}{n} \theta. \end{aligned}$$

(c) If $C = (\emptyset, E)$, then

$$\begin{aligned} \phi_i^C(L, R, \sigma_{-i}) &= p(\mathcal{X}_{-i}^\sigma = 0) \left[u(|t_{LE}^w - t_i|) - u(|t_R - t_i|) - \bar{d} + \left(s_E + \frac{1}{n} \right) \theta \right] \\ &\quad + p(\mathcal{X}_{-i}^\sigma = 2e) [u(|t_L - t_i|) - u(|t_{LE}^w - t_i|) + \bar{d}] - \frac{1}{n} \theta. \end{aligned}$$

³⁹Note that the lemma only considers pure strategies $\sigma'_i = L$ and $\sigma''_i = R$. In the rare cases where mixed strategies are required for the proofs of the results presented in the rest of the paper, we leave the calculation of ϕ_i^C to the proof itself.

(d) If $C = (E, E)$, then

$$\begin{aligned}
\phi_i^C(L, R, \sigma_{-i}) &= p(\mathcal{X}_{-i}^\sigma = 0) \left[u \left(\left| t \left(\gamma + \frac{1}{2n}, \frac{1}{2} - \frac{1}{2n} \right) - t_i \right| \right) \right. \\
&\quad \left. - u(|t_R - t_i|) - \theta \left(\gamma - \frac{1}{2n} \right) \right] \\
&+ \sum_{j=1}^{2e-1} p(\mathcal{X}_{-i}^\sigma = j) \left[u \left(\left| t \left(\gamma + \frac{1}{2n} + \frac{j}{n}, \frac{1}{2} - \frac{1}{2n} - \frac{j}{n} \right) - t_i \right| \right) \right. \\
&\quad \left. - u \left(\left| t \left(\gamma - \frac{1}{2n} + \frac{j}{n}, \frac{1}{2} + \frac{1}{2n} - \frac{j}{n} \right) - t_i \right| \right) \right] \\
&+ p(\mathcal{X}_{-i}^\sigma = 2e) \left[u(|t_L - t_i|) - u \left(\left| t \left(\frac{1}{2} - \frac{1}{2n}, \gamma + \frac{1}{2n} \right) - t_i \right| \right) \right. \\
&\quad \left. + \theta \left(\gamma - \frac{1}{2n} \right) \right].
\end{aligned}$$

Proof of Lemma 1 (Polarization)

Let $j \in \{w, s\}$ be arbitrary but fixed in the following arguments.

(A) Case: $i \in \Omega^R \cup \{\frac{1}{2}\}$, that is $t_i \in [\frac{1}{2}, t_R^\Omega]$.

Firstly, consider the set of inequalities with a positive “+” sign, that is (i) is the version with $\bar{d} \geq s_E \theta + \left(u(|t_{LE}^j - t_i|) - u(|t_R - t_i|) \right)$ for example. Then, we have that (ii) \implies (i) and (iv) \implies (iii), since $u(t_i - t_L) \leq u(|t_R - t_i|) \forall t_i \in [\frac{1}{2}, t_R^\Omega]$. Moreover, (iv) \implies (ii), because $u(|t_{RE}^j - t_i|) \geq u(t_i - t_{LE}^j)$. Thus, we need only show (iv). For this, we show that $u(|t_{RE}^j - t_i|) - u(t_i - t_L)$ is increasing in t_i on $[\frac{1}{2}, t_R^\Omega]$. Indeed, if $t_i < t_{RE}^j$, we have

$$\frac{\partial}{\partial t_i} \left\{ u(|t_{RE}^j - t_i|) - u(t_i - t_L) \right\} = -u'(t_{RE}^j - t_i) - u'(t_i - t_L) > 0$$

since $u'(x) < 0 \forall x > 0$ is assumed. On the other hand, if $t_i > t_{RE}^j$, we have

$$\frac{\partial}{\partial t_i} \left\{ u(|t_{RE}^j - t_i|) - u(t_i - t_L) \right\} = u'(t_{RE}^j - t_i) - u'(t_i - t_L) \geq 0$$

because u' is decreasing (u is assumed to be (weakly) concave).⁴⁰ Now the fact that $u(|t_{RE}^j - t_i|) - u(t_i - t_L)$ is increasing in t_i implies that

$$u(|t_{RE}^j - t_R^\Omega|) - u(t_R^\Omega - t_L) \geq u(|t_{RE}^j - t_R^\Omega|) - u(t_R^\Omega - t_L) \quad \forall t_i \in \left[\frac{1}{2}, t_R^\Omega\right]$$

So that (iv) is implied by

$$\bar{d} \geq s_E \theta + u(|t_{RE}^j - t_R^\Omega|) - u(t_R^\Omega - t_L)$$

which are precisely the PCs.

Secondly, consider the set of inequalities with a negative “-” sign, that is (i) is the version with $\bar{d} \geq s_E \theta - \left(u(|t_{LE}^j - t_i|) - u(|t_R - t_i|)\right)$ for example. Then, by the same arguments as above, we have that (i) \implies (ii), (iii) \implies (iv), and (i) \implies (iii). Thus, we need only show (i). For this, we show that $-u(|t_{LE}^j - t_i|) + u(|t_R - t_i|)$ is increasing in t_i on $[\frac{1}{2}, t_R^\Omega]$. Indeed, if $t_i < t_R$, we have

$$\frac{\partial}{\partial t_i} \left\{ u(|t_R - t_i|) - u(t_i - t_{LE}^j) \right\} = -u'(t_R - t_i) - u'(t_i - t_{LE}^j) > 0$$

since $u'(x) < 0 \quad \forall x > 0$ is assumed. On the other hand, if $t_i > t_R$, we have

$$\frac{\partial}{\partial t_i} \left\{ u(|t_R - t_i|) - u(t_i - t_{LE}^j) \right\} = u'(t_i - t_R) - u'(t_i - t_{LE}^j) \geq 0$$

because u' is (weakly) decreasing. Now as before, the fact that $u(|t_R - t_i|) - u(t_i - t_{LE}^j)$ is increasing in t_i implies that

$$u(|t_R - t_R^\Omega|) - u(t_R^\Omega - t_{LE}^j) = u(|\eta|) - u(\pi + \eta - \beta^j) \geq u(|t_R - t_i|) - u(t_i - t_{LE}^j) \quad \forall t_i \in \left[\frac{1}{2}, t_R^\Omega\right]$$

so that (i) is implied by

$$\bar{d} \geq s_E \theta + u(|\eta|) - u(\pi + \eta - \beta^j)$$

⁴⁰Technically, $u(x)$ is not differentiable in the classic two-sided limit sense, at $x = 0$. Thus, we should treat the cases $t_i = t_L = t_{RE}^j = \frac{1}{2}$ (recall that inequality (1) is assumed) and $t_i = t_{RE}^j$ separately. However, the fact that $u(|t_{RE}^j - t_i|) - u(t_i - t_L)$ is increasing on $[\frac{1}{2}, t_R^\Omega]$ follows from the positive derivative via the mean-value theorem, which for the case $t_{RE}^j \neq \frac{1}{2}$ would have to be applied twice, once on the subdomain $[\frac{1}{2}, t_{RE}^j]$ and once on the subdomain $[t_{RE}^j, t_R^\Omega]$. The same issues arise when we make the same monotonicity arguments for the other cases below. We leave it to the reader to apply the mean-value theorem in the correct way.

We claim that this is weaker condition than the PCs. We have that $|\eta| \leq |\beta^j + \eta|$, but also that $0 \leq \pi + \eta - \beta^j \leq \pi + \eta$. Thus this is a not entirely trivial claim. Let $g : [0, 1] \rightarrow \mathbb{R}$ be the following function:

$$g(\alpha) = u(|\alpha\beta^j + \eta|) - u(\pi + \eta + (\alpha - 1)\beta^j).$$

Then $\alpha = 1$ corresponds to the quantity which appears in the PCs, and $\alpha = 0$ corresponds to the quantity appearing above. We show that g is increasing in α . Indeed, first assume that $\eta \geq 0$. Then

$$g'(\alpha) = \beta^j u'(\alpha\beta^j + \eta) - \beta^j u'(\pi + \eta + (\alpha - 1)\beta^j)$$

so that since $u'(x)$ is (weakly) decreasing in x ,

$$g'(\alpha) \geq 0 \quad \Leftrightarrow \quad \pi + \eta + (\alpha - 1)\beta^j \geq \alpha\beta^j + \eta \quad \forall \alpha \in [0, 1].$$

But as this is equivalent to $\pi - \beta^j \geq 0$, the claim is proved by the mean-value theorem.

Now assume that $\eta < 0$. Then we must distinguish the cases $\alpha < \bar{\alpha} = -\frac{\eta}{\beta^j}$ and $\alpha \geq \bar{\alpha}$. If $\alpha \geq \bar{\alpha}$, $\alpha\beta^j + \eta \geq 0$ so the derivative is the same as above. If $\alpha < \bar{\alpha}$, we have

$$g'(\alpha) = -\beta^j u'(-\alpha\beta^j - \eta) - \beta^j u'(\pi + \eta + (\alpha - 1)\beta^j) > 0$$

since $u'(x) < 0 \quad \forall x > 0$, and the claim is proved again by the mean-value theorem.

(B) Case: $i \in \Omega^L$, that is $t_i \in [t_L^\Omega, \frac{1}{2}]$.

Firstly, consider again the set of inequalities with a positive “+” sign. Then, by the same arguments as before, we have that *mutatis mutandis*, $(i) \implies (ii)$, $(iii) \implies (iv)$, and $(i) \implies (iii)$. Thus, we need only show (i) . Analogously to before, we show that $u(|t_{LE}^j - t_i|) - u(t_R - t_i)$ is decreasing in t_i on $[t_L^\Omega, \frac{1}{2}]$. Indeed, if $t_i < t_{LE}^j$, we have

$$\frac{\partial}{\partial t_i} \left\{ u(|t_{LE}^j - t_i|) - u(t_R - t_i) \right\} = -u'(t_{LE}^j - t_i) + u'(t_R - t_i) \leq 0$$

since u' is decreasing (u is assumed to be (weakly) concave). On the other hand, if $t_i > t_{LE}^j$, we have

$$\frac{\partial}{\partial t_i} \left\{ u(|t_{LE}^j - t_i|) - u(t_R - t_i) \right\} = u'(t_i - t_{LE}^j) + u'(t_R - t_i) < 0$$

because $u'(x) < 0 \forall x > 0$ is assumed. Now the fact that $u(|t_{LE}^j - t_i|) - u(t_R - t_i)$ is decreasing in t_i implies that

$$u(|t_{LE}^j - t_L^\Omega|) - u(t_R - t_L^\Omega) \geq u(|t_{LE}^j - t_i|) - u(t_R - t_i) \quad \forall t_i \in \left[t_L^\Omega, \frac{1}{2} \right)$$

which in turn shows that the PCs imply (i).

Secondly, consider the set of inequalities with a negative “-” sign. Then, we have that (ii) \implies (i), (iv) \implies (iii), and (iv) \implies (ii). Thus, we need only show (iv). As in the “+” sign case above, one sees that $u(|t_L - t_i|) - u(t_{RE}^j - t_i)$ is decreasing in t_i on $\left[t_L^\Omega, \frac{1}{2} \right)$. This implies that it is sufficient to show

$$\bar{d} \geq s_E \theta + u(|t_L - t_L^\Omega|) - u(t_{RE}^j - t_L^\Omega) = s_E \theta + u(|\eta|) - u(\pi + \eta - \beta^j).$$

But in the $i \in \Omega^R \cup \left\{ \frac{1}{2} \right\}$ above, we showed that this is implied by the PCs.

Proof of Proposition 1 (Nash equilibria of \mathcal{G})

We distinguish four cases depending on the CPPs written in Stage 1.

Case (a): $C = (\emptyset, \emptyset)$.

First, we prove that σ^{LR} is an equilibrium. Indeed, according to Assumption 2, if voters vote according to σ^{LR} , the implemented policy is either (t_{LE}^s, \bar{d}) or (t_{RE}^s, \bar{d}) with probability $\frac{1}{2}$, and support $s_G = \frac{1}{2}(1 - s_E) + \frac{1}{2n}$ in both cases. If $i \in \Omega^L$ deviates and votes strategically for R , the implemented policy is (t_{RE}^s, \bar{d}) for sure with support $s_G \pm \frac{1}{n}$, while if $i \in \Omega^R$ deviates and votes strategically for L , the implemented policy is (t_{LE}^s, \bar{d}) for sure with support $s_G \pm \frac{1}{n}$. Then, for each $i \in \Omega^L$,

$$\begin{aligned} & \mathbb{E} \left[U_i \left((L, \sigma_{-i}^{LR}) \right) \right] - \mathbb{E} \left[U_i \left((R, \sigma_{-i}^{LR}) \right) \right] \\ &= \left[\frac{u(|t_i - t_{LE}^s|) + u(|t_i - t_{RE}^s|)}{2} - \bar{d} - \theta s_G \right] - \left[u(|t_i - t_{RE}^s|) - \bar{d} - \theta \left(s_G + \frac{1}{2n} \right) \right] \\ &= \frac{u(|t_i - t_{LE}^s|) - u(|t_i - t_{RE}^s|)}{2} + \frac{\theta}{2n} > 0, \end{aligned}$$

since by assumption u is decreasing, $t_i < \frac{1}{2}$ implies $|t_i - t_{LE}^s| \leq t_{RE}^s - t_i$ and $\theta > 0$.

Analogously, for each $i \in \Omega^R$, we obtain

$$\mathbb{E} \left[U_i \left((R, \sigma_{-i}^{LR}) \right) \right] - \mathbb{E} \left[U_i \left((L, \sigma_{-i}^{LR}) \right) \right] > 0.$$

Lastly, by symmetry, the median voter, m , is indifferent between voting for either party, i.e.:

$$\mathbb{E} \left[U_i \left(\left(\frac{1}{2} LR, \sigma_{-i}^{LR} \right) \right) \right] = \mathbb{E} \left[U_i \left((L, \sigma_{-i}^{LR}) \right) \right] = \mathbb{E} \left[U_i \left((L, \sigma_{-i}^{LR}) \right) \right].$$

As a consequence, σ^{LR} is a (weak⁴¹) Nash equilibrium of the voting subgame.

Second, we prove that σ^L is an equilibrium if the Strong PC holds. According to Assumption 2, if voters vote according to σ^L , the implemented policy is $(t_L, 0)$ with support $s_G = \frac{1}{2} + \frac{1}{2n}$. If $i \in \Omega$ deviates and votes for R , the implemented policy is (t_{LE}^s, \bar{d}) with support $s_G = \frac{1}{2} - \frac{1}{2n}$. Then, for each $i \in \Omega$,

$$\begin{aligned} & \mathbb{E} \left[U_i \left((L, \sigma_{-i}^L) \right) \right] - \mathbb{E} \left[U_i \left((R, \sigma_{-i}^L) \right) \right] \\ &= \left[u(|t_i - t_L|) - \theta \left(\frac{1}{2} + \frac{1}{2n} \right) \right] - \left[u(|t_i - t_{LE}^s|) - \bar{d} - \theta \left(\frac{1}{2} - \frac{1}{2n} \right) \right] \\ &= u(|t_i - t_L|) - u(|t_i - t_{LE}^s|) + \bar{d} - \frac{\theta}{n}. \end{aligned}$$

We see that if

$$\bar{d} \geq \frac{\theta}{n} + u(|t_i - t_{LE}^s|) - u(|t_i - t_L|) \quad \forall i \in \Omega, \quad (16)$$

then it is not profitable for any $i \in \Omega$ to deviate and σ^L is a Nash equilibrium. However, since $s_E > 0$ and hence $e \geq 1$ is assumed, we have $s_E = \frac{2e}{n} > \frac{1}{n}$, and the fact that the above inequality holds strictly follows from Lemma 1.(ii).⁴² Hence σ^L is a strict Nash equilibrium.

Third, we focus on σ^R . According to Assumption 2, if voters vote according to σ^R , the implemented policy is $(t_R, 0)$ with support $s_G = \frac{1}{2} + \frac{1}{2n}$. If $i \in \Omega$ deviates and votes for L , the implemented policy is (t_{RE}^s, \bar{d}) with support $s_G = \frac{1}{2} - \frac{1}{2n}$. Analogously to the analysis for σ^L , we can show that σ^R is an equilibrium if

$$\bar{d} \geq \frac{\theta}{n} + u(|t_i - t_{RE}^s|) - u(|t_i - t_R|) \quad \forall i \in \Omega$$

which is again implied by Lemma 1.(iii), and as before, the equilibrium is strict.

Case (b): $C = (E, \emptyset)$.

⁴¹Because of the median voter.

⁴²It can be shown in a similar fashion to the proof of Lemma 1 that $u(|t_i - t_{LE}^s|) - u(|t_i - t_L|)$ is increasing in t_i for all $t_i \in [t_L^\Omega, t_R^\Omega]$. Hence, one can show that $\exists n \in \mathbb{N}$ such that σ^L is a Nash equilibrium if and only if $\bar{d} > u(|t_R^\Omega - t_{LE}^s|) - u(|t_R^\Omega - t_L|) = u(\pi + \eta - \beta^s) - u(\pi + \eta)$.

First, we prove that σ^{LR} is not an equilibrium. Indeed, according to Assumption 2, if voters vote according to σ^{LR} , the implemented policy is (t_{RE}^w, \bar{d}) for sure with support $s_G = \frac{1-s_E}{2} \pm \frac{1}{2n}$.⁴³ If $i \in \Omega^R$ deviates and votes strategically for L , the implemented policy is (t_{RE}^w, \bar{d}) for sure with support $s_G = \frac{1-s_E}{2} - \frac{1}{n} \pm \frac{1}{2n}$. In particular, for each $i \in \Omega^R$,

$$\begin{aligned} & \mathbb{E} \left[U_i \left((R, \sigma_{-i}^{LR}) \right) \right] - \mathbb{E} \left[U_i \left((L, \sigma_{-i}^{LR}) \right) \right] \\ &= \left[u(|t_i - t_{RE}^w|) - \bar{d} - \theta \left(\frac{1-s_E}{2} \right) \right] - \left[u(|t_i - t_{RE}^w|) - \bar{d} - \theta \left(\frac{1-s_E}{2} - \frac{1}{n} \right) \right] \\ &= -\frac{1}{n} \theta < 0. \end{aligned}$$

Hence, σ^{LR} is not a Nash equilibrium of the voting subgame.

Second, we prove that σ^L is an equilibrium if the Weak PC holds. Indeed, according to Assumption 2, if voters vote according to σ^L , the implemented policy is $(t_L, 0)$ with support $s_G = \frac{1}{2} + \frac{1}{2n}$. If $i \in \Omega$ deviates and votes for R , the implemented policy is (t_{RE}^w, \bar{d}) with support $s_G = \frac{1}{2} - s_E + \frac{1}{2n}$. We stress that the coalition between E and L is now ruled out by the CPPs written in the previous stage. Then, for each $i \in \Omega$,

$$\begin{aligned} & \mathbb{E} \left[U_i \left((L, \sigma_{-i}^L) \right) \right] - \mathbb{E} \left[U_i \left((R, \sigma_{-i}^L) \right) \right] \\ &= \left[u(|t_i - t_L|) - \theta \left(\frac{1}{2} + \frac{1}{2n} \right) \right] - \left[u(|t_i - t_{RE}^w|) - \bar{d} - \theta \left(\frac{1}{2} - s_E + \frac{1}{2n} \right) \right] \\ &= u(|t_i - t_L|) - u(|t_i - t_{RE}^w|) + \bar{d} - \theta s_E \geq 0, \end{aligned}$$

by Lemma 1.(iv). Note that we need this inequality to hold also for $t_i = t_R^\Omega$, and thus, σ^L is a strict Nash equilibrium if and only if the Weak PC holds strictly.

Third, we study when σ^R is an equilibrium. According to Assumption 2, if voters vote according to σ^R , the implemented policy is $(t_R, 0)$ with support $s_G = \frac{1}{2} + \frac{1}{2n}$. If $i \in \Omega$ deviates and votes for L , the implemented policy is (t_{RE}^w, \bar{d}) with support $s_G = \frac{1}{2} - \frac{1}{2n}$. Analogously to the σ^L case, it can be checked that σ^R is an equilibrium of the voting subgame in Case (b) if and only if

$$\bar{d} \geq \frac{\theta}{n} + u(|t_i - t_{RE}^w|) - u(|t_i - t_R|) \quad \forall i \in \Omega.$$

But the strict version of this inequality follows immediately from Lemma 1.(iii) and the fact that $e \geq 1$ is assumed.

⁴³Recall that in the sincere voting equilibrium the median voter votes for R and L with probability 1/2 and therefore $s_G = \frac{1-s_E}{2} + \frac{1}{2n}$ if they voted for R and $s_G = \frac{1-s_E}{2} - \frac{1}{2n}$ if they voted for L .

Case (c): $C = (\emptyset, E)$.

This case can be analyzed in an analogous way to Case (b), to obtain that σ^{LR} is not an equilibrium; that σ^L is an equilibrium if

$$\bar{d} \geq \frac{\theta}{n} + u(|t_i - t_{LE}^w|) - u(|t_i - t_L|) \quad \forall i \in \Omega, \quad (17)$$

which again holds strictly (and hence σ^L is a strict equilibrium) as long as the Weak PC holds; and that σ^R is an equilibrium if

$$\bar{d} \geq s_E \theta + u(|t_i - t_{LE}^w|) - u(|t_i - t_R|) \quad \forall i \in \Omega, \quad (18)$$

which is implied by Lemma 1.(i) and holds strictly for every i if and only if the Weak PC holds strictly.

Case (d): $C = (E, E)$.

First, we prove that σ^{LR} is an equilibrium. Indeed, let $s^{LR} = \frac{1-s_E-\frac{1}{n}}{2}$, and recall the definition of

$$t(s_L, s_R) = \operatorname{argmax}_{t \in [t_L, t_R]} s_L u(t - t_L) + s_R u(t_R - t).$$

Since our model assumptions mean that we always have $s_L + s_R = 1 - s_E$, for notational convenience we suppress the dependence of $t(s_L, s_R)$ on s_R for the rest of this proof, that is we let

$$t(s_L) := \underbrace{t(s_L, 1 - s_E - s_L)}_{\text{original definition}}.$$

According to Assumption 2, if voters vote according to σ^{LR} , the implemented policies are $(t(s^{LR} + \frac{1}{n}), 0)$ or $(t(s^{LR}), 0)$, each with probability $\frac{1}{2}$ and support $s_G = 1 - s_E$. If $i \in \Omega^L$ deviates and votes strategically for R , the implemented policies are $(t(s^{LR}), 0)$ or $(t(s^{LR} - \frac{1}{n}), 0)$, each with probability $\frac{1}{2}$ and support $s_G = 1 - s_E$. If $i \in \Omega^R$ deviates and votes strategically for L , the implemented policies are $(t(s^{LR} + \frac{2}{n}), 0)$ or $(t(s^{LR} + \frac{1}{n}), 0)$, each with probability $\frac{1}{2}$ and support $s_G = 1 - s_E$. Note that since t_L and t_R are assumed to be symmetric around $\frac{1}{2}$ and u is assumed to

be strictly decreasing, if $t_L \neq t_R$ ⁴⁴, we have:

$$t(s_L) \text{ is strictly decreasing in } s_L \quad (19)$$

$$t\left(s^{LR} + \frac{2}{n}\right) + t\left(s^{LR} - \frac{1}{n}\right) = 1$$

$$t\left(s^{LR} + \frac{1}{n}\right) + t(s^{LR}) = 1 \quad (20)$$

Then, for each $i \in \Omega^L$,

$$\begin{aligned} & \mathbb{E} \left[U_i \left((L, \sigma_{-i}^{LR}) \right) \right] - \mathbb{E} \left[U_i \left((R, \sigma_{-i}^{LR}) \right) \right] \\ &= \left[\frac{u\left(|t_i - t\left(s^{LR} + \frac{1}{n}\right)|\right) + u\left(|t_i - t(s^{LR})|\right)}{2} - \theta s_G \right] \\ & \quad - \left[\frac{u\left(|t_i - t(s^{LR})|\right) + u\left(|t_i - t\left(s^{LR} - \frac{1}{n}\right)|\right)}{2} - \theta s_G \right] \\ &= \frac{u\left(|t_i - t\left(s^{LR} + \frac{1}{n}\right)|\right) - u\left(|t_i - t\left(s^{LR} - \frac{1}{n}\right)|\right)}{2} > 0, \end{aligned}$$

because $t_i < \frac{1}{2}$ and $t\left(s^{LR} + \frac{1}{n}\right) < t\left(s^{LR} - \frac{1}{n}\right)$, by (19).

Analogously, for each $i \in \Omega^R$, we obtain

$$\mathbb{E} \left[U_i \left((R, \sigma_{-i}^{LR}) \right) \right] - \mathbb{E} \left[U_i \left((L, \sigma_{-i}^{LR}) \right) \right] > 0$$

Lastly, by symmetry (that is, equation (20)), the median voter, m , is indifferent between voting for either party. As a consequence, σ^{LR} is a (weak) Nash equilibrium of the voting subgame.

Second, we consider σ^L and show that this is not an equilibrium if and only if the SPC holds. Indeed, according to Assumption 2, if voters vote according to σ^L , the implemented policy is $(t_L, 0)$ with support $s_G = \frac{1}{2} + \frac{1}{2n}$. If $i \in \Omega$ deviates and votes for R , the implemented policy is $\left(t\left(s_G - \frac{1}{n}\right), 0\right)$ with support $s_G^i = 1 - s_E$. Note that $t_L < t\left(s_G - \frac{1}{n}\right)$. Then, for each $i \in \Omega$,

$$\begin{aligned} & \mathbb{E} \left[U_i \left((L, \sigma_{-i}^L) \right) \right] - \mathbb{E} \left[U_i \left((R, \sigma_{-i}^L) \right) \right] \\ &= [u(|t_i - t_L|) - \theta s_G] - \left[u\left(\left|t_i - t\left(s_G - \frac{1}{n}\right)\right|\right) - \theta(1 - s_E) \right] \\ &= u\left(|t_i - t_L|\right) - u\left(\left|t_i - t\left(s_G - \frac{1}{n}\right)\right|\right) + \theta\left(\frac{1}{2} - s_E - \frac{1}{2n}\right) \end{aligned}$$

⁴⁴If $t_L = t_R = \frac{1}{2}$, then of course $t(s_L, s_R) = t_m = \frac{1}{2}$ whatever the vote shares s_L and s_R .

By the usual argument, we show that this is strictly decreasing in t_i . We distinguish three cases. Firstly, assume that $t_i < t_L$. Then,

$$\frac{\partial}{\partial t_i} \left\{ u(|t_i - t_L|) - u \left(\left| t_i - t \left(s_G - \frac{1}{n} \right) \right| \right) \right\} = -u'(t_L - t_i) + u' \left(t \left(s_G - \frac{1}{n} \right) - t_i \right) \leq 0,$$

since $u'(\cdot)$ is decreasing. Secondly, assume that $t_L < t_i < t \left(s_G - \frac{1}{n} \right)$. Then,

$$\frac{\partial}{\partial t_i} \left\{ u(|t_i - t_L|) - u \left(\left| t_i - t \left(s_G - \frac{1}{n} \right) \right| \right) \right\} = u'(t_i - t_L) - u' \left(t \left(s_G - \frac{1}{n} \right) - t_i \right) < 0,$$

because $u'(x) < 0 \forall x > 0$ is assumed. Finally assume that $t_i > t \left(s_G - \frac{1}{n} \right)$. Then,

$$\frac{\partial}{\partial t_i} \left\{ u(|t_i - t_L|) - u \left(\left| t_i - t \left(s_G - \frac{1}{n} \right) \right| \right) \right\} = u'(t_i - t_L) - u' \left(t_i - t \left(s_G - \frac{1}{n} \right) t_L \right) \leq 0,$$

where again the inequality holds since $u(\cdot)$ is (weakly) concave. As a result, by the mean-value theorem, we have:

$$\begin{aligned} & u(|t_i - t_L|) - u \left(\left| t_i - t \left(s_G - \frac{1}{n} \right) \right| \right) + \theta \left(\frac{1}{2} - s_E - \frac{1}{2n} \right) \\ & \geq u(|t_R^\Omega - t_L|) - u \left(\left| t_R^\Omega - t \left(s_G - \frac{1}{n} \right) \right| \right) + \theta \left(\frac{1}{2} - s_E - \frac{1}{2n} \right). \end{aligned}$$

Since the right-hand side of this inequality is attained when $t_i = t_R^\Omega$, this implies that σ^L is an equilibrium if and only if

$$\theta \left(\gamma - \frac{1}{2n} \right) \geq u \left(\left| t_R^\Omega - t \left(s_G - \frac{1}{n} \right) \right| \right) - u(|t_R^\Omega - t_L|), \quad (21)$$

which is precisely the inverse inequality appearing in the SPC.

Third, and finally, analogously to the σ^L case, one shows that σ^R is an equilibrium if and only if

$$\theta \left(\gamma - \frac{1}{2n} \right) \geq u \left(\left| t_L^\Omega - t \left(\gamma + \frac{1}{2n} \right) \right| \right) - u(|t_L^\Omega - t_R|). \quad (22)$$

Because of the symmetry of t_L and t_R around $\frac{1}{2}$, we have $t \left(s_G - \frac{1}{n} \right) + t \left(\gamma + \frac{1}{2n} \right) = 1$ and thus

$$\left| t_L^\Omega - t \left(\gamma + \frac{1}{2n} \right) \right| = \left| t_R^\Omega - t \left(s_G - \frac{1}{n} \right) \right|$$

so that the right-hand side of the inequality in (22) is identical to the right-hand side of (21) and both are precisely the inverse of the SPC.

Proof of Proposition 2

The perk relevant government seat shares, s_G , for the different pairs of equilibrium strategies and governments are given by:

$$\begin{aligned} s_{(\sigma^L, L)} &= s_{(\sigma^R, R)} = \frac{1}{2} + \frac{1}{2n} \\ s_{(\sigma^{LR}, LR)} &= 1 - s_E \\ s_{(\sigma^{LR}, LE)} &= s_{(\sigma^{LR}, RE)} = \frac{1}{2}(1 - s_E) + \frac{1}{2n}. \end{aligned}$$

Proof (i).

By the formula in (7), we have:

$$\begin{aligned} W\left(t_L, 0, \frac{1}{2} + \frac{1}{2n}\right) &= \bar{u}(t_L) - \theta\left(\frac{1}{2} + \frac{1}{2n}\right) \\ W\left(t_R, 0, \frac{1}{2} + \frac{1}{2n}\right) &= \bar{u}(t_R) - \theta\left(\frac{1}{2} + \frac{1}{2n}\right). \end{aligned}$$

By symmetry of all voters around the median, these two quantities are identical. On the other hand,

$$\begin{aligned} W(t_{LR}^L, 0, 1 - s_E) &\stackrel{\text{symmetry}}{=} W(t_{LR}^R, 0, 1 - s_E) \\ &= \bar{u}(t_{LR}^L) - \theta(1 - s_E), \end{aligned}$$

and we see that

$$\begin{aligned} W(\Phi(\sigma^{LR}, LR, t_{LR}^L)) &= W(t_{LR}^L, 0, 1 - s_E) = W(\Phi(\sigma^{LR}, LR, t_{LR}^R)) \\ &> W(\Phi(\sigma^L, L)) = W\left(t_L, 0, \frac{1}{2} + \frac{1}{2n}\right) = W(\Phi(\sigma^R, R)) \\ \Leftrightarrow \theta\left(\gamma - \frac{1}{2n}\right) &< \bar{u}(t_{LR}^L) - \bar{u}(t_L), \end{aligned} \tag{23}$$

where Φ is a well-defined function which maps any triple of equilibrium strategy profile, resulting government and policy to its corresponding triple (t, d, s_G) . We claim that if n is large enough, WSPC implies (23), while \neg WSPC holding strictly, that is,

$$\theta\gamma > \bar{u}(t_m) - \bar{u}(t_L),$$

implies the strict reverse inequality. Indeed, even if both

$$\begin{aligned} \theta\gamma &> \theta\left(\gamma - \frac{1}{2n}\right), \text{ and} \\ \bar{u}(t_m) - \bar{u}(t_L) &> \bar{u}(t_{LR}^L) - \bar{u}(t_L), \end{aligned}$$

by the fact that the argmax of the Nash-bargaining solution in the case of a grand coalition is unique for our party utilities and that the Maximum Theorem in Beavis and Dobbs (1990) implies that $t_{LR}^L(n)$ is continuous in n ,

$$\begin{aligned} \lim_{n \rightarrow \infty} t_{LR}^L(n) &= \lim_{n \rightarrow \infty} t \left(\frac{1-s_E}{2} + \frac{1}{2n}, \frac{1-s_E}{2} - \frac{1}{2n} \right) = t_m = \frac{1}{2}, \text{ and} \\ \lim_{n \rightarrow \infty} \left\{ \gamma - \frac{1}{2n} \right\} &= \gamma. \end{aligned}$$

If n is large enough, the left-hand and right-hand sides of (23) can both be made to be contained in disjoint ε -balls of $\theta\gamma$ and $\bar{u}(t_m) - \bar{u}(t_L)$ respectively. This proves the first statement in (i).

Proof (ii).

Let $j \in \{w, s\}$ indicate the extreme party's bargaining power. We have:

$$\begin{aligned} W \left(t_{LE}^j, \bar{d}, \frac{1-s_E}{2} + \frac{1}{2n} \right) &= W \left(t_{RE}^j, \bar{d}, \frac{1-s_E}{2} + \frac{1}{2n} \right) \\ &= \bar{u}(t_{LE}^j) - 2\gamma\bar{d} - \frac{\theta}{2} \left(1 - s_E + \frac{1}{n} \right), \end{aligned}$$

and we see that this is less than or equal to $W(t_{LR}^L, 0, 1 - s_E)$ if and only if

$$2\gamma\bar{d} \geq \frac{\theta}{2} \left(1 - s_E - \frac{1}{n} \right) + \bar{u}(t_{LE}^j) - \bar{u}(t_{LR}^L). \quad (24)$$

However, as long as $t_{LE}^s < \frac{1}{2}$,⁴⁵ in the limit

$$\bar{u}(t_{LE}^s) - \lim_{n \rightarrow \infty} \bar{u}(t_{LR}^L(n)) < 0,$$

and so inequality (24) will be implied by the WC if n is large enough. $t_{LE}^s = \frac{1}{2}$ is equivalent to $t_L = t_R$, which implies that $t_{LR}^L(n) = \frac{1}{2}$ for all n , so that the welfare ordering in (ii) is true for all n .

Proof of Lemma 2

Let σ be an arbitrary strategy profile. We distinguish four cases depending on the CPPs written in Stage 1. Further we define for each $i \in \Omega$ and $\sigma = (q_i)_{i \in \Omega}$, the random variable \mathcal{X}_{-i}^σ that counts the number of votes for party L in $\Omega \setminus \{i\}$, given that players

⁴⁵Recall that $t_{LE}^w \leq t_{LE}^s$, by assumption.

different than i play according to σ_{-i} . We build on Assumption 2 and the discussion in subsection 4.1 to identify the government formation process – and the corresponding policy decision – after each election outcome, depending on the realised value of the number of left-wing votes amongst players other than i , χ_{-i}^σ .

Case (a): $C = (\emptyset, \emptyset)$.

On the one hand,

$$\begin{aligned} \mathbb{E} [U_i((L, \sigma_{-i}))] &= p(\mathcal{X}_{-i}^\sigma = 0) \left[u(|t_{RE}^s - t_i|) - \bar{d} - \theta \left(\frac{1}{2} - \frac{1}{2n} \right) \right] \\ &+ \sum_{j=1}^{e-1} p(\mathcal{X}_{-i}^\sigma = j) \left[u(|t_{RE}^s - t_i|) - \bar{d} - \theta \left(\frac{1}{2} - \frac{1}{2n} - \frac{j}{n} \right) \right] \\ &+ p(\mathcal{X}_{-i}^\sigma = e) \left[u(|t_{LE}^s - t_i|) - \bar{d} - \theta \left(\frac{1}{2}(1 - s_E) + \frac{1}{2n} \right) \right] \\ &+ \sum_{j=e+1}^{2e-1} p(\mathcal{X}_{-i}^\sigma = j) \left[u(|t_{LE}^s - t_i|) - \bar{d} - \theta \left(\frac{1}{2}(1 - s_E) + \frac{1}{2n} + \frac{j-e}{n} \right) \right] \\ &+ p(\mathcal{X}_{-i}^\sigma = 2e) \left[u(|t_L - t_i|) - \theta \left(\frac{1}{2} + \frac{1}{2n} \right) \right]. \end{aligned}$$

On the other hand,

$$\begin{aligned} \mathbb{E} [U_i((R, \sigma_{-i}))] &= p(\mathcal{X}_{-i}^\sigma = 0) \left[u(|t_R - t_i|) - \theta \left(\frac{1}{2} + \frac{1}{2n} \right) \right] \\ &+ \sum_{j=1}^{e-1} p(\mathcal{X}_{-i}^\sigma = j) \left[u(|t_{RE}^s - t_i|) - \bar{d} - \theta \left(\frac{1}{2} + \frac{1}{2n} - \frac{j}{n} \right) \right] \\ &+ p(\mathcal{X}_{-i}^\sigma = e) \left[u(|t_{RE}^s - t_i|) - \bar{d} - \theta \left(\frac{1}{2}(1 - s_E) + \frac{1}{2n} \right) \right] \\ &+ \sum_{j=e+1}^{2e-1} p(\mathcal{X}_{-i}^\sigma = j) \left[u(|t_{LE}^s - t_i|) - \bar{d} - \theta \left(\frac{1}{2}(1 - s_E) - \frac{1}{2n} + \frac{j-e}{n} \right) \right] \\ &+ p(\mathcal{X}_{-i}^\sigma = 2e) \left[u(|t_{LE}^s - t_i|) - \bar{d} - \theta \left(\frac{1}{2} - \frac{1}{2n} \right) \right]. \end{aligned}$$

Therefore,

$$\begin{aligned} &\mathbb{E} [U_i((L, \sigma_{-i}))] - \mathbb{E} [U_i((R, \sigma_{-i}))] \\ &= \frac{1}{n} \theta \left[\sum_{j=0}^{e-1} p(\mathcal{X}_{-i}^\sigma = j) - \sum_{j=e+1}^{2e} p(\mathcal{X}_{-i}^\sigma = j) \right] + \bar{d} [p(\mathcal{X}_{-i}^\sigma = 2e) - p(\mathcal{X}_{-i}^\sigma = 0)] \\ &+ p(\mathcal{X}_{-i}^\sigma = 0) [u(|t_{RE}^s - t_i|) - u(|t_R - t_i|)] + p(\mathcal{X}_{-i}^\sigma = 2e) [u(|t_L - t_i|) - u(|t_{LE}^s - t_i|)] \\ &+ p(\mathcal{X}_{-i}^\sigma = e) [u(|t_{LE}^s - t_i|) - u(|t_{RE}^s - t_i|)]. \end{aligned}$$

Case (b): $C = (E, \emptyset)$.

On the one hand,

$$\begin{aligned}\mathbb{E} [U_i((L, \sigma_{-i}))] &= p(\mathcal{X}_{-i}^\sigma = 0) \left[u(|t_{RE}^w - t_i|) - \bar{d} - \theta \left(\frac{1}{2} - \frac{1}{2n} \right) \right] \\ &\quad + \sum_{j=1}^{2e-1} p(\mathcal{X}_{-i}^\sigma = j) \left[u(|t_{RE}^w - t_i|) - \bar{d} - \theta \left(\frac{1}{2} - \frac{1}{2n} - \frac{j}{n} \right) \right] \\ &\quad + p(\mathcal{X}_{-i}^\sigma = 2e) \left[u(|t_L - t_i|) - \theta \left(\frac{1}{2} + \frac{1}{2n} \right) \right].\end{aligned}$$

On the other hand,

$$\begin{aligned}\mathbb{E} [U_i((R, \sigma_{-i}))] &= p(\mathcal{X}_{-i}^\sigma = 0) \left[u(|t_R - t_i|) - \theta \left(\frac{1}{2} + \frac{1}{2n} \right) \right] \\ &\quad + \sum_{j=1}^{2e-1} p(\mathcal{X}_{-i}^\sigma = j) \left[u(|t_{RE}^w - t_i|) - \bar{d} - \theta \left(\frac{1}{2} + \frac{1}{2n} - \frac{j}{n} \right) \right] \\ &\quad + p(\mathcal{X}_{-i}^\sigma = 2e) \left[u(|t_{RE}^w - t_i|) - \bar{d} - \theta \left(\frac{1}{2} - s_E + \frac{1}{2n} \right) \right].\end{aligned}$$

Therefore,

$$\begin{aligned}&\mathbb{E} [U_i((L, \sigma_{-i}))] - \mathbb{E} [U_i((R, \sigma_{-i}))] \\ &= p(\mathcal{X}_{-i}^\sigma = 0) \left[u(|t_{RE}^w - t_i|) - u(|t_R - t_i|) - \bar{d} + \frac{1}{n}\theta \right] + \sum_{j=1}^{2e-1} p(\mathcal{X}_{-i}^\sigma = j) \frac{1}{n}\theta \\ &\quad + p(\mathcal{X}_{-i}^\sigma = 2e) [u(|t_L - t_i|) - u(|t_{RE}^w - t_i|) + \bar{d} - s_E\theta] \\ &= p(\mathcal{X}_{-i}^\sigma = 0) [u(|t_{RE}^w - t_i|) - u(|t_R - t_i|) - \bar{d}] \\ &\quad + p(\mathcal{X}_{-i}^\sigma = 2e) \left[u(|t_L - t_i|) - u(|t_{RE}^w - t_i|) + \bar{d} - \left(s_E + \frac{1}{n} \right) \theta \right] + \frac{1}{n}\theta.\end{aligned}$$

Case (c): $C = (\emptyset, E)$.

On the one hand,

$$\begin{aligned}\mathbb{E} [U_i((L, \sigma_{-i}))] &= p(\mathcal{X}_{-i}^\sigma = 0) \left[u(|t_{LE}^w - t_i|) - \bar{d} - \theta \left(\frac{1}{2} - s_E + \frac{1}{2n} \right) \right] \\ &\quad + \sum_{j=1}^{2e-1} p(\mathcal{X}_{-i}^\sigma = j) \left[u(|t_{LE}^w - t_i|) - \bar{d} - \theta \left(\frac{1}{2} - s_E + \frac{1}{2n} + \frac{j}{n} \right) \right] \\ &\quad + p(\mathcal{X}_{-i}^\sigma = 2e) \left[u(|t_L - t_i|) - \theta \left(\frac{1}{2} + \frac{1}{2n} \right) \right].\end{aligned}$$

On the other hand,

$$\begin{aligned}
\mathbb{E} [U_i((R, \sigma_{-i}))] &= p(\mathcal{X}_{-i}^\sigma = 0) \left[u(|t_R - t_i|) - \theta \left(\frac{1}{2} + \frac{1}{2n} \right) \right] \\
&\quad + \sum_{j=1}^{2e-1} p(\mathcal{X}_{-i}^\sigma = j) \left[u(|t_{LE}^w - t_i|) - \bar{d} - \theta \left(\frac{1}{2} - s_E - \frac{1}{2n} + \frac{j}{n} \right) \right] \\
&\quad + p(\mathcal{X}_{-i}^\sigma = 2e) \left[u(|t_{LE}^w - t_i|) - \bar{d} - \theta \left(\frac{1}{2} - \frac{1}{2n} \right) \right].
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathbb{E} [U_i((L, \sigma_{-i}))] - \mathbb{E} [U_i((R, \sigma_{-i}))] \\
&= p(\mathcal{X}_{-i}^\sigma = 0) \left[u(|t_{LE}^w - t_i|) - u(|t_R - t_i|) - \bar{d} + s_E \theta \right] - \sum_{j=1}^{2e-1} p(\mathcal{X}_{-i}^\sigma = j) \frac{1}{n} \theta \\
&\quad + p(\mathcal{X}_{-i}^\sigma = 2e) \left[u(|t_L - t_i|) - u(|t_{LE}^w - t_i|) + \bar{d} - \frac{1}{n} \theta \right] \\
&= p(\mathcal{X}_{-i}^\sigma = 0) \left[u(|t_{LE}^w - t_i|) - u(|t_R - t_i|) - \bar{d} + \left(s_E + \frac{1}{n} \right) \theta \right] \\
&\quad + p(\mathcal{X}_{-i}^\sigma = 2e) \left[u(|t_L - t_i|) - u(|t_{LE}^w - t_i|) + \bar{d} \right] - \frac{1}{n} \theta.
\end{aligned}$$

Case (d): $C = (E, E)$.

On the one hand,

$$\begin{aligned}
&\mathbb{E} [U_i((L, \sigma_{-i}))] \\
&= p(\mathcal{X}_{-i}^\sigma = 0) \left[u \left(\left| t \left(\frac{1}{2} - s_E + \frac{1}{2n}, \frac{1}{2} - \frac{1}{2n} \right) - t_i \right| \right) - \theta(1 - s_E) \right] \\
&\quad + \sum_{j=1}^{2e-1} p(\mathcal{X}_{-i}^\sigma = j) \left[u \left(\left| t \left(\frac{1}{2} - s_E + \frac{1}{2n} + \frac{j}{n}, \frac{1}{2} - \frac{1}{2n} - \frac{j}{n} \right) - t_i \right| \right) - \theta(1 - s_E) \right] \\
&\quad + p(\mathcal{X}_{-i}^\sigma = 2e) \left[u(|t_L - t_i|) - \theta \left(\frac{1}{2} + \frac{1}{2n} \right) \right].
\end{aligned}$$

On the other hand,

$$\begin{aligned}
& \mathbb{E} [U_i((R, \sigma_{-i}))] \\
&= p(\mathcal{X}_{-i}^\sigma = 0) \left[u(|t_R - t_i|) - \theta \left(\frac{1}{2} + \frac{1}{2n} \right) \right] \\
&+ \sum_{j=1}^{2e-1} p(\mathcal{X}_{-i}^\sigma = j) \left[u \left(\left| t \left(\frac{1}{2} - s_E - \frac{1}{2n} + \frac{j}{n}, \frac{1}{2} + \frac{1}{2n} - \frac{j}{n} \right) - t_i \right| \right) - \theta(1 - s_E) \right] \\
&+ p(\mathcal{X}_{-i}^\sigma = 2e) \left[u \left(\left| t \left(\frac{1}{2} - \frac{1}{2n}, \frac{1}{2} - s_E + \frac{1}{2n} \right) - t_i \right| \right) - \theta(1 - s_E) \right].
\end{aligned}$$

Therefore,

$$\begin{aligned}
& \mathbb{E} [U_i((L, \sigma_{-i}))] - \mathbb{E} [U_i((R, \sigma_{-i}))] \\
&= p(\mathcal{X}_{-i}^\sigma = 0) \left[u \left(\left| t \left(\gamma + \frac{1}{2n}, \frac{1}{2} - \frac{1}{2n} \right) - t_i \right| \right) - u(|t_R - t_i|) - \theta \left(\gamma - \frac{1}{2n} \right) \right] \\
&+ \sum_{j=1}^{2e-1} p(\mathcal{X}_{-i}^\sigma = j) \left[u \left(\left| t \left(\gamma + \frac{1}{2n} + \frac{j}{n}, \frac{1}{2} - \frac{1}{2n} - \frac{j}{n} \right) - t_i \right| \right) \right. \\
&\quad \left. - u \left(\left| t \left(\gamma - \frac{1}{2n} + \frac{j}{n}, \frac{1}{2} + \frac{1}{2n} - \frac{j}{n} \right) - t_i \right| \right) \right] \\
&+ p(\mathcal{X}_{-i}^\sigma = 2e) \left[u(|t_L - t_i|) - u \left(\left| t \left(\frac{1}{2} - \frac{1}{2n}, \gamma + \frac{1}{2n} \right) - t_i \right| \right) + \theta \left(\gamma - \frac{1}{2n} \right) \right].
\end{aligned}$$

Proof of Lemma 3

In this section, we provide and prove the following Lemma to establish continuity of $\phi_i(\sigma'_i, \sigma''_i, (1 - \varepsilon)\sigma'_{-i} + \varepsilon\sigma''_{-i})$ in ε .

Lemma 3

Suppose we have a game with player set \mathcal{N} , arbitrary strategy sets S_i (which could be uncountable) for each player $i \in \mathcal{N}$, and non- i strategy sets $S_{-i} = \prod_{j \in \mathcal{N} \setminus \{i\}} S_j$. Let $\sigma'_i, \sigma''_i : X \rightarrow S_i$ and $\sigma'_{-i}, \sigma''_{-i} : X \rightarrow S_{-i}$ be random variables on some arbitrary probability space (X, \mathcal{F}, P) (to account for mixed strategies). For a subset of players $I \subset \mathcal{N} \setminus \{i\}$, let $(I\sigma'_{-i}, I^c\sigma''_{-i}) : X \rightarrow S_{-i}$ denote the mixed strategy resulting from all players in I playing σ' and all players in $I^c = \mathcal{N} \setminus (I \cup \{i\})$ playing σ'' . Suppose that randomization about which strategy σ' or σ'' is chosen by each player $j \in \mathcal{N} \setminus \{i\}$, as the result of their playing the mixed strategy $(1 - \varepsilon)\sigma'_j + \varepsilon\sigma''_j$, is independent of all

other random variables in the game, that the player set is finite, $|\mathcal{N}| < \infty$, and that the random variables $U_i(\sigma', (I\sigma'_{-i}, I^c\sigma''_{-i}))$, $U_i(\sigma'', (I\sigma'_{-i}, I^c\sigma''_{-i}))$ are (X, \mathcal{F}, P) -integrable (i.e. expectations exist) for every subset $I \subset \mathcal{N} \setminus \{i\}$. Then, for fixed σ' , σ'' , and $i \in \mathcal{N}$, the function

$$\begin{aligned} \phi_i(\sigma'_i, \sigma''_i, (1-\varepsilon)\sigma'_{-i} + \varepsilon\sigma''_{-i}) \\ := \mathbb{E} [U_i(\sigma'_i, (1-\varepsilon)\sigma'_{-i} + \varepsilon\sigma''_{-i}) - U_i(\sigma''_i, (1-\varepsilon)\sigma'_{-i} + \varepsilon\sigma''_{-i})] \end{aligned}$$

is a polynomial of order $|\mathcal{N}| - 1$ in ε . In particular, it is continuous, has at most $|\mathcal{N}| - 1$ roots, and

$$\begin{aligned} &\exists B \subset \mathbb{R} \text{ with non-zero Lebesgue measure such that} \\ &\phi_i(\sigma'_i, \sigma''_i, (1-\varepsilon)\sigma'_{-i} + \varepsilon\sigma''_{-i}) = 0 \quad \forall \varepsilon \in B \\ \Leftrightarrow &\sigma'_i = \sigma''_i. \end{aligned}$$

Proof.

Essentially this follows from the fact for two independent random variables Y and Z , $\mathbb{E}[YZ] = \mathbb{E}[Y]\mathbb{E}[Z]$. Indeed, we may write:

$$\sigma_{-i} = (1-\varepsilon)\sigma'_{-i} + \varepsilon\sigma''_{-i} = \sum_{x=0}^{|\mathcal{N}|-1} \sum_{I \subset \mathcal{N} \setminus \{i\} : |I|=x} (I\sigma'_{-i}, I^c\sigma''_{-i}) 1_{\mathcal{I}=I},$$

where recall, $(I\sigma'_{-i}, I^c\sigma''_{-i})$ refers to the random variable $X \rightarrow S_{-i}$ such that $j \in I$ plays σ'_j and $j \in I^c$ plays σ''_j and such that the components are assumed to be in the right (player) order, and where \mathcal{I} is the random variable $X \rightarrow \mathcal{P}(\mathcal{N} \setminus \{i\})$ which chooses the set of players which play σ' as the result of the $(1-\varepsilon)/\varepsilon$ randomization process. Note that $\mathcal{P}(\mathcal{N} \setminus \{i\})$ refers to the power set of $\mathcal{N} \setminus \{i\}$. In turn, the expression above immediately implies that

$$\begin{aligned} \phi_i(\sigma'_i, \sigma''_i, (1-\varepsilon)\sigma'_{-i} + \varepsilon\sigma''_{-i}) \\ = \sum_{x=0}^{|\mathcal{N}|-1} \sum_{I \subset \mathcal{N} \setminus \{i\} : |I|=x} \mathbb{E} \left[\left\{ U_i(\sigma'_i, (I\sigma'_{-i}, I^c\sigma''_{-i})) - U_i(\sigma''_i, (I\sigma'_{-i}, I^c\sigma''_{-i})) \right\} 1_{\mathcal{I}=I} \right] \\ = \sum_{x=0}^{|\mathcal{N}|-1} \sum_{I \subset \mathcal{N} \setminus \{i\} : |I|=x} \mathbb{E} \left[\left\{ U_i(\sigma'_i, (I\sigma'_{-i}, I^c\sigma''_{-i})) - U_i(\sigma''_i, (I\sigma'_{-i}, I^c\sigma''_{-i})) \right\} \right] \mathbb{E} [1_{\mathcal{I}=I}] \end{aligned}$$

since by assumption \mathcal{I} is independent of $\sigma'_i, \sigma''_i, \sigma'_{-i}$, etc. However,

$$\mathbb{E} [1_{\mathcal{I}=I}] = P(\mathcal{I} = I) = \binom{|\mathcal{N}| - 1}{|I|} \varepsilon^{|I|} (1-\varepsilon)^{|\mathcal{N}|-|I|-1},$$

and so the above can be written as

$$\begin{aligned} & \phi_i(\sigma'_i, \sigma''_i, (1 - \varepsilon)\sigma'_{-i} + \varepsilon\sigma''_{-i}) \\ &= \sum_{x=0}^{|\mathcal{N}|-1} P\left(B_{1-\varepsilon}^{|\mathcal{N}|-1} = x\right) \\ & \quad \sum_{I \subset \mathcal{N} \setminus \{i\} : |I|=x} \mathbb{E} \left[\left\{ U_i(\sigma'_i, (I\sigma'_{-i}, I^c\sigma''_{-i})) - U_i(\sigma''_i, (I\sigma'_{-i}, I^c\sigma''_{-i})) \right\} \right], \end{aligned}$$

where $B_{1-\varepsilon}^{|\mathcal{N}|-1}$ is a binomially distributed random variable with number of trials $|\mathcal{N}| - 1$ and success probability $1 - \varepsilon$. Now for each x , σ' , and σ'' , the second sum is a constant, while for each x , $P\left(B_{1-\varepsilon}^{|\mathcal{N}|-1}\right)$ is a polynomial of degree $|\mathcal{N}| - 1$ in ε . Thus, $\phi_i(\sigma'_i, \sigma''_i, (1 - \varepsilon)\sigma'_{-i} + \varepsilon\sigma''_{-i})$ must also be a polynomial of degree at most $|\mathcal{N}| - 1$ in ε .

C.2 Proof of Proposition 3

(iii) \Leftrightarrow (iv).

This is obvious, both for the weak and strict inequality cases.

(i) \implies (iii).

Consider the function $g_i(\varepsilon) := (1 - \varepsilon)\phi_i^C(\sigma''_i, \sigma'_i, \sigma''_{-i}) + \varepsilon\phi_i^C(\sigma'_i, \sigma'_i, \sigma'_{-i})$. Then, by definition of the Nash equilibrium (equation (??)) and equation (10), we have

$$\begin{aligned} \phi_i^C(\sigma''_i, \sigma'_i, \sigma''_{-i}) &\geq 0 \text{ and} \\ \phi_i^C(\sigma'_i, \sigma'_i, \sigma'_{-i}) &\leq 0, \end{aligned}$$

which means that $g_i(\varepsilon)$ is a linearly decreasing function going through 0 once at

$$\varepsilon = \varepsilon^0 := \frac{\phi_i^C(\sigma''_i, \sigma'_i, \sigma''_{-i})}{\phi_i^C(\sigma''_i, \sigma'_i, \sigma''_{-i}) - \phi_i^C(\sigma'_i, \sigma'_i, \sigma'_{-i})} = \frac{\phi_i^C(\sigma''_i, \sigma'_i, \sigma''_{-i})}{\phi_i^C(\sigma''_i, \sigma'_i, \sigma''_{-i}) + \phi_i^C(\sigma'_i, \sigma'_i, \sigma'_{-i})},$$

and thus the maximum level at which σ'' is immune to correlated errors w.r.t. σ' for player i is precisely ε^0 . In turn, using antisymmetry (equation (10)), this implies that

$$\sigma' \succsim_i^{\parallel} \sigma'' \quad (\text{respectively } \succ_i^{\parallel} \text{ for the strict domination case}) \quad (25)$$

$$\Leftrightarrow \forall \varepsilon \leq \varepsilon^0 \quad (1 - \varepsilon)\phi_i^C(\sigma'_i, \sigma''_i, \sigma'_{-i}) - \varepsilon\phi_i^C(\sigma''_i, \sigma'_i, \sigma''_{-i}) \geq 0 \quad (\text{respectively } > 0) \quad (26)$$

Suppose now toward a contradiction that $\phi_i^C(\sigma''_i, \sigma'_i, \sigma''_{-i}) > \phi_i^C(\sigma'_i, \sigma'_i, \sigma'_{-i})$ (respectively \geq). Then $\varepsilon^0 > 1 - \varepsilon^0$ (respectively \geq). But then the left-hand side of inequality (26)

must be < 0 (respectively ≤ 0). Contradiction to weak (respectively strict) domination $\sigma' \succ_i^{\parallel} \sigma''$.

(iii) \implies (i).

We know that $\phi_i^C(\sigma'_i, \sigma''_i, \sigma'_{-i}) \geq \phi_i^C(\sigma''_i, \sigma'_i, \sigma''_{-i})$ (respectively $\phi_i^C(\sigma'_i, \sigma''_i, \sigma'_{-i}) > \phi_i^C(\sigma''_i, \sigma'_i, \sigma''_{-i})$). This immediately implies

$$1 - \varepsilon^0 \geq \varepsilon^0 \quad (\text{respectively } 1 - \varepsilon^0 > \varepsilon^0),$$

so that inequality (26) (respectively the strict version) holds.

(iii) \implies (ii).

Consider the function $h_i(\varepsilon) := (1-\varepsilon)\phi_i^C(\sigma'_i, \sigma''_i, \sigma'_{-i}) + \varepsilon\phi_i^C(\sigma'_i, \sigma''_i, \sigma''_{-i}) = (1-\varepsilon)\phi_i^C(\sigma'_i, \sigma''_i, \sigma'_{-i}) - \varepsilon\phi_i^C(\sigma''_i, \sigma'_i, \sigma''_{-i})$, which is the relevant quantity in the definition of σ' being immune to correlated errors w.r.t. σ'' . h_i is almost identical to g_i , except that it has its unique 0 at $1 - \varepsilon^0$, rather than ε^0 . If $\phi_i^C(\sigma'_i, \sigma''_i, \sigma'_{-i}) \geq \phi_i^C(\sigma''_i, \sigma'_i, \sigma''_{-i})$, $1 - \varepsilon^0 \geq \frac{1}{2} \geq \varepsilon^0$, which implies that $h_i(\varepsilon) \geq 0 \forall \varepsilon \leq \frac{1}{2}$. If $\phi_i^C(\sigma'_i, \sigma''_i, \sigma'_{-i}) \geq \phi_i^C(\sigma''_i, \sigma'_i, \sigma''_{-i})$, then $1 - \varepsilon^0 > \frac{1}{2} > \varepsilon^0$ and so indeed $\delta = 1 - \varepsilon^0 - \frac{1}{2}$.

(ii) \implies (iii).

Again, suppose towards a contradiction that $\phi_i^C(\sigma''_i, \sigma'_i, \sigma''_{-i}) > \phi_i^C(\sigma'_i, \sigma''_i, \sigma'_{-i})$. Then $\varepsilon^0 > \frac{1}{2} > 1 - \varepsilon^0$ and $h_i(\varepsilon) < 0 \forall \varepsilon > 1 - \varepsilon^0$ which is in contradiction to immunity at level $\frac{1}{2}$. For the strict inequality case, suppose that $\phi_i^C(\sigma''_i, \sigma'_i, \sigma''_{-i}) \geq \phi_i^C(\sigma'_i, \sigma''_i, \sigma'_{-i})$. Then $\varepsilon^0 \geq \frac{1}{2} \geq 1 - \varepsilon^0$ so that $h_i(\varepsilon) < 0 \forall \varepsilon > \frac{1}{2}$, which is in contradiction to immunity at level $\frac{1}{2} + \delta$ for some δ .

C.3 Proof of Proposition 4

Let $f(\varepsilon) := f_i(\varepsilon)$ be defined as in the statement of the Proposition, where for notational convenience we now drop the subscript i . The key to this Proposition is recognizing that by antisymmetry of ϕ in its first two arguments (equation (10)), we have

$$\phi_i^C(\sigma''_i, \sigma'_i, (1-\varepsilon)\sigma''_{-i} + \varepsilon\sigma'_{-i}) = -f(1-\varepsilon),$$

where the latter function is the relevant quantity appearing in the definition of σ'' being immune to uncorrelated errors w.r.t. σ' for player i .

(i).

By definition, σ' being immune to uncorrelated errors at level $\varepsilon^* = \frac{1}{2}$ w.r.t. σ'' for player i means that

$$f(\varepsilon) \geq 0 \quad \forall \varepsilon \in \left[0, \frac{1}{2}\right]. \quad (27)$$

But this immediately implies that $-f(1 - \varepsilon) \leq 0 \quad \forall \varepsilon \in [\frac{1}{2}, 1]$. By the remark above, this already proves $\sigma' \succsim_i^\perp \sigma''$, provided that we show that the pathological case, where $\exists \varepsilon^* > \frac{1}{2}$ such that $-f(1 - \varepsilon) = 0 \quad \forall \varepsilon \in [\frac{1}{2}, \varepsilon^*]$ while $f(\varepsilon) \leq 0 \quad \forall \varepsilon \in [\frac{1}{2}, 1]$ and $\exists \varepsilon \in [\frac{1}{2}, 1]$ such that $f(\varepsilon) < 0$ (hence $-f(1 - \varepsilon) \geq 0 \quad \forall \varepsilon \in [0, \frac{1}{2}]$), cannot occur. In that case, σ'' would be immune to uncorrelated errors w.r.t. σ' at level ε^* , but not the other way around. But this impossibility follows from the fact that $-f(1 - \varepsilon)$ is a polynomial with discrete roots however. Indeed, unless $\sigma'_i = \sigma''_i$, every left-neighbourhood $[\frac{1}{2} - \delta, \frac{1}{2}]$ ($\delta > 0$) of $\frac{1}{2}$ must contain a point ε where $f(\varepsilon) > 0$ (recall that (27) holds). But then also every right neighbourhood $[\frac{1}{2}, \frac{1}{2} + \delta]$ must contain a point ε where $-f(1 - \varepsilon) < 0$, and thus σ'' can be immune to uncorrelated at level at most $\frac{1}{2}$ w.r.t. σ' for i . The case $\sigma'_i = \sigma''_i$ is not very interesting, as then $f(\varepsilon) = -f(1 - \varepsilon) = 0 \quad \forall \varepsilon \in [0, 1]$, and $\sigma' \sim_i^\perp \sigma''$.

For the strict domination case, note that if $f(\varepsilon) \geq 0 \quad \forall \varepsilon \in [0, \frac{1}{2} + \delta]$, $\exists \varepsilon^0 \in (\frac{1}{2}, \frac{1}{2} + \delta)$ such that $f(\varepsilon^0) > 0$. In turn this implies of course that $-f(1 - (1 - \varepsilon^0)) < 0$. Thus, σ'' can be immune to uncorrelated errors at level at most $\varepsilon^* < 1 - \varepsilon^0 < \frac{1}{2}$ w.r.t. σ' and $\sigma'' \not\succeq_i^\perp \sigma'$. Also note that by continuity,

$$\begin{aligned} & \exists \delta > 0 \text{ such that } \sigma' \text{ is immune to uncorrelated errors at level } \frac{1}{2} + \delta \text{ w.r.t. } \sigma'' \\ \Leftrightarrow & f(\varepsilon) \geq 0 \quad \forall \varepsilon \in \left[0, \frac{1}{2}\right] \text{ and } f\left(\frac{1}{2}\right) > 0. \end{aligned}$$

(ii).

$\sigma' \succsim_i^\perp \sigma''$, we must have $f(\frac{1}{2}) \geq 0$. Indeed, if $f(\frac{1}{2}) < 0$, since $f(\varepsilon) \leq f(\frac{1}{2}) < 0 \quad \forall \varepsilon \in [\frac{1}{2}, 1]$, by assumption, we would have $-f(1 - \varepsilon) > 0 \quad \forall \varepsilon \in [0, \frac{1}{2}]$ and σ'' would be immune to uncorrelated errors at level $\frac{1}{2}$ w.r.t. σ' , but not the other way around. But then also $f(\varepsilon) \geq f(\frac{1}{2}) \geq 0 \quad \forall \varepsilon \in [0, \frac{1}{2}]$ so σ' is immune to uncorrelated errors at level $\frac{1}{2}$ w.r.t. σ'' .

For the strict case, note that with our “near-monotonicity” property, if $\sigma' \succsim_i^\perp \sigma''$,

$f\left(\frac{1}{2}\right) = 0$ is no longer possible, as this would imply

$$f(\varepsilon) \geq 0 \quad \Leftrightarrow \quad -f(1 - \varepsilon) \geq 0 \quad \Leftrightarrow \quad \varepsilon \in \left[0, \frac{1}{2}\right]$$

from which it would follow that both σ' and σ'' are immune w.r.t. each other at maximal level exactly $\frac{1}{2}$ and so $\sigma'_i \sim_i^\perp \sigma''_i$. As noted above, by continuity, $f(\varepsilon) \geq 0 \quad \forall \varepsilon \in \left[0, \frac{1}{2}\right]$ and $f\left(\frac{1}{2}\right) > 0$ is equivalent to immunity of σ' w.r.t. σ'' at level $\frac{1}{2} + \delta$ for some $\delta > 0$.

(iii).

Let $f(\varepsilon) := f_i(\varepsilon)$, as in the statement of the Proposition, where for convenience we drop the subscript i . Analogously to the previous parts of the proof, we use the fact that

$$\phi_i^C(\sigma''_i, \sigma'_i, (1 - \varepsilon)\sigma''_{-i} + \varepsilon\sigma'_{-i}) = -f_i(1 - \varepsilon) = -f(1 - \varepsilon).$$

$f(\varepsilon) + f(1 - \varepsilon) \geq 0 \quad \forall \varepsilon \in [0, 1]$ is equivalent to $f(\varepsilon) \geq -f(1 - \varepsilon) \quad \forall \varepsilon \in [0, 1]$. Thus if $-f(1 - \varepsilon) \geq 0 \quad \forall \varepsilon \in [0, \varepsilon^*]$, then obviously also $f(\varepsilon) \geq 0 \quad \forall \varepsilon \in [0, \varepsilon^*]$.

For the strict domination case, let $\varepsilon_{\sigma'', \sigma'}^0$ be the maximum level at which σ'' is immune to uncorrelated errors w.r.t. σ' for i . In other words, $\varepsilon_{\sigma'', \sigma'}^0$ is the smallest root of $-f(1 - \varepsilon)$ in $[0, 1]$. By assumption, we must have $f(\varepsilon_{\sigma'', \sigma'}^0) > -f(1 - \varepsilon_{\sigma'', \sigma'}^0) = 0$. But then by continuity and the fact that $f(\varepsilon) > -f(1 - \varepsilon) \quad \forall \varepsilon \in [0, 1]$, the smallest root of f in $[0, 1]$, call it $\varepsilon_{\sigma', \sigma''}^0$, must be strictly greater than $\varepsilon_{\sigma'', \sigma'}^0$. This means that σ' is immune at level $\varepsilon_{\sigma', \sigma''}^0$, w.r.t. σ'' , but not the other way around, so that $\sigma'' \not\sim_i^\perp \sigma'$.

C.4 Proof of Lemma 4

In this section, we formulate and prove the following lemma.

Lemma 4

Let $C = (\emptyset, \emptyset)$, $\sigma' = \sigma^L$, $\sigma'' = \sigma^R$, $i \in \Omega^R$, and B_p^l be a binomial random variable with number of trials l and success probability p that is, $B_p^l \sim \text{Bin}(l, p)$. Then:

$$\begin{aligned} \phi_i^{(\emptyset, \emptyset)}(\sigma_i^L, \sigma_i^R, (1 - \varepsilon)\sigma_{-i}^L + \varepsilon\sigma_{-i}^R) &= \frac{\theta}{n} \left[2 \sum_{j=0}^{e-1} p(B_{1-\varepsilon}^{2e} = j) - 1 \right] \\ &+ \binom{2e}{e} (1 - \varepsilon)^e \varepsilon^e \left[u(t_i - t_{LE}^s) - u(|t_{RE}^s - t_i|) + \frac{\theta}{n} \right] \\ &+ \varepsilon^{2e} \left[u(|t_{RE}^s - t_i|) - u(|t_R - t_i|) - \bar{d} \right] \\ &+ (1 - \varepsilon)^{2e} \left[u(t_i - t_L) - u(t_i - t_{LE}^s) + \bar{d} \right]. \end{aligned}$$

Define the three following quantities, which depend on the model parameters t_i , $u(\cdot)$, t_L , t_R , t_{LE}^s and t_{RE}^s , but not on ε :

$$\begin{aligned} K_1 &:= u(t_i - t_{LE}^s) - u(|t_{RE}^s - t_i|), \\ K_2 &:= u(|t_{RE}^s - t_i|) - u(|t_R - t_i|) - \bar{d}, \\ K_3 &:= u(t_i - t_L) - u(t_i - t_{LE}^s) + \bar{d}. \end{aligned}$$

Note that by our assumptions about the policies, $u(\cdot)$ and the strong PC (Lemma 1), for $i \in \Omega^R$, we must necessarily have $K_1 < 0$, $K_2 < 0$, and $K_3 > 0$. Moreover, the proof of the Lemma in the Appendix shows that $K_3 < |K_2|$.

Then, $\sigma^R \succ_i^{\frac{1}{2}} \sigma^L$, for all $i \in \Omega^R$ and all values of K_1 , K_2 , K_3 , n , e and θ . In addition, we provide four examples of parameter vectors $(n, e, \theta, K_1, K_2, K_3)$ where σ^R is not immune to uncorrelated errors at level $\frac{1}{2}$ w.r.t. σ^L .

Lemma 4 proves that no exact equivalent of Proposition 3 can hold for the uncorrelated case. We also mention that we did not specifically search for the parameter values used in the Lemma, such that σ^R is not immune to uncorrelated errors at level $\frac{1}{2}$ w.r.t. σ^L . In fact, for three of them we used one of the most simple $u(\cdot)$ satisfying our assumptions, namely $u(x) = -\frac{1}{2}x^2$. In that sense, we should view the counterexample given in the Lemma as rather natural, and not some specifically designed pathological case.

Proof.

We have $\chi_i^{(1-\varepsilon)\sigma^L + \varepsilon\sigma^R} = B_{1-\varepsilon}^{2e}$. Thus, according to Lemma 2, we calculate:

$$\begin{aligned} \phi_i^C(\sigma_i^L, \sigma_i^R, (1-\varepsilon)\sigma_{-i}^L + \varepsilon\sigma_{-i}^R) &=: f_i(\varepsilon) = \\ \frac{\theta}{n} \left[\sum_{j=0}^{e-1} p(B_{1-\varepsilon}^{2e} = j) - \sum_{j=e+1}^{2e} p(B_{1-\varepsilon}^{2e} = j) \right] &+ \binom{2e}{e} (1-\varepsilon)^e \varepsilon^e [u(t_i - t_{LE}^s) - u(|t_{RE}^s - t_i|)] \\ &+ \varepsilon^{2e} [u(|t_{RE}^s - t_i|) - u(|t_R - t_i|) - \bar{d}] \\ &+ (1-\varepsilon)^{2e} [u(t_i - t_L) - u(t_i - t_{LE}^s) + \bar{d}], \end{aligned}$$

and after rewriting

$$- \sum_{j=e+1}^{2e} p(B_{1-\varepsilon}^{2e} = j) = \sum_{j=0}^e p(B_{1-\varepsilon}^{2e} = j) - 1,$$

we have the expression in the statement of the Lemma.

We proceed to show that $f_i(\varepsilon) + f_i(1-\varepsilon) < 0 \forall \varepsilon \in [0, 1] \forall i \in \Omega^R$, which by Proposition 4 (iii) shows $\sigma^R \succ_i^{\frac{1}{2}} \sigma^L \forall i \in \Omega^R$. (Recall that $\phi_i^C(\sigma_i^R, \sigma_i^L, (1-\varepsilon)\sigma_{-i}^R + \varepsilon\sigma_{-i}^L) = -f_i(1-\varepsilon)$

ε .) The key to this is to show:

$$\sum_{j=0}^e p(B_{1-\varepsilon}^{2e} = j) = \sum_{\tilde{j}=e+1}^{2e} p(B_{\varepsilon}^{2e} = \tilde{j}).$$

However, this is intuitively obvious. Formally, consider the general term in the first sum:

$$\binom{2e}{j} (1-\varepsilon)^j \varepsilon^{2e-j}.$$

since $\binom{2e}{j} = \binom{2e}{2e-j}$, this term appears exactly once when $\tilde{j} = 2e - j$, that is the first term $j = 0$ of the first sum is equal to the last term $\tilde{j} = 2e$ of the second sum, etc.

With this result, we now have

$$\begin{aligned} & f_i(\varepsilon) + f_i(1-\varepsilon) \\ &= \frac{\theta}{n} \left[2 \sum_{j=0}^{e-1} p(B_{1-\varepsilon}^{2e} = j) - 1 + 2 \sum_{j=0}^{e-1} p(B_{\varepsilon}^{2e} = j) - 1 \right] \\ & \quad + 2 \binom{2e}{e} (1-\varepsilon)^e \varepsilon^e \left[u(t_i - t_{LE}^s) - u(|t_{RE}^s - t_i|) + \frac{\theta}{n} \right] \\ & \quad + \{ \varepsilon^{2e} + (1-\varepsilon)^{2e} \} [u(|t_{RE}^s - t_i|) - u(|t_R - t_i|) - \bar{d}] \\ & \quad + \{ (1-\varepsilon)^{2e} + \varepsilon^{2e} \} [u(t_i - t_L) - u(t_i - t_{LE}^s) + \bar{d}] \\ &= \frac{\theta}{n} [2(1 - p(B_{\varepsilon}^{2e} = e)) - 2] + \frac{\theta}{n} 2p(B_{1-\varepsilon}^{2e} = e) + 2p(B_{1-\varepsilon}^{2e} = e)K_1 \\ & \quad + \{ \varepsilon^{2e} + (1-\varepsilon)^{2e} \} [K_2 + K_3] \\ &= 2p(B_{1-\varepsilon}^{2e} = e)K_1 + \{ \varepsilon^{2e} + (1-\varepsilon)^{2e} \} [K_2 + K_3] \end{aligned}$$

The first term is < 0 for $i \in \Omega^R$. We now proceed to show that $K_2 + K_3 < 0$, or in other words, as stated in the statement of the Lemma, that $K_3 < |K_2|$. Indeed:

$$K_2 + K_3 = u(|t_{RE}^s - t_i|) - u(|t_R - t_i|) + u(t_i - t_L) - u(t_i - t_{LE}^s).$$

For $t_i \geq t_R$, both $K_2 < 0$ and $K_3 < 0$, so this case is obvious. For the other cases, we show that the above is decreasing as a function of t_i for $t_i \in (\frac{1}{2}, t_R)$. For $t_i \in (\frac{1}{2}, t_{RE}^s)$:

$$\begin{aligned} & \frac{\partial}{\partial t_i} \{ u(|t_{RE}^s - t_i|) - u(|t_R - t_i|) + u(t_i - t_L) - u(t_i - t_{LE}^s) \} \\ &= -u'(t_i - t_{RE}^s) + u'(t_R - t_i) + u'(t_i - t_L) - u'(t_i - t_{LE}^s) \leq 0, \end{aligned}$$

since $u'(x)$ is assumed to be weakly decreasing in x . For $t_i \in (t_{RE}^s, t_R)$:

$$\begin{aligned} & \frac{\partial}{\partial t_i} \{ u(|t_{RE}^s - t_i|) - u(|t_R - t_i|) + u(t_i - t_L) - u(t_i - t_{LE}^s) \} \\ &= u'(t_i - t_{RE}^s) + u'(t_R - t_i) + u'(t_i - t_L) - u'(t_i - t_{LE}^s) \leq 0, \end{aligned}$$

since $u'(x) < 0 \forall x > 0$ and $u'(x)$ is assumed to be weakly decreasing. But by symmetry of the policies around $\frac{1}{2}$, when $t_i = \frac{1}{2}$, $K_2 + K_3 = 0$. Thus $K_2 + K_3 \leq 0$ for all $i \in \Omega^R$. In turn, this gives us the required $f_i(\varepsilon) + f_i(1 - \varepsilon) < 0$.

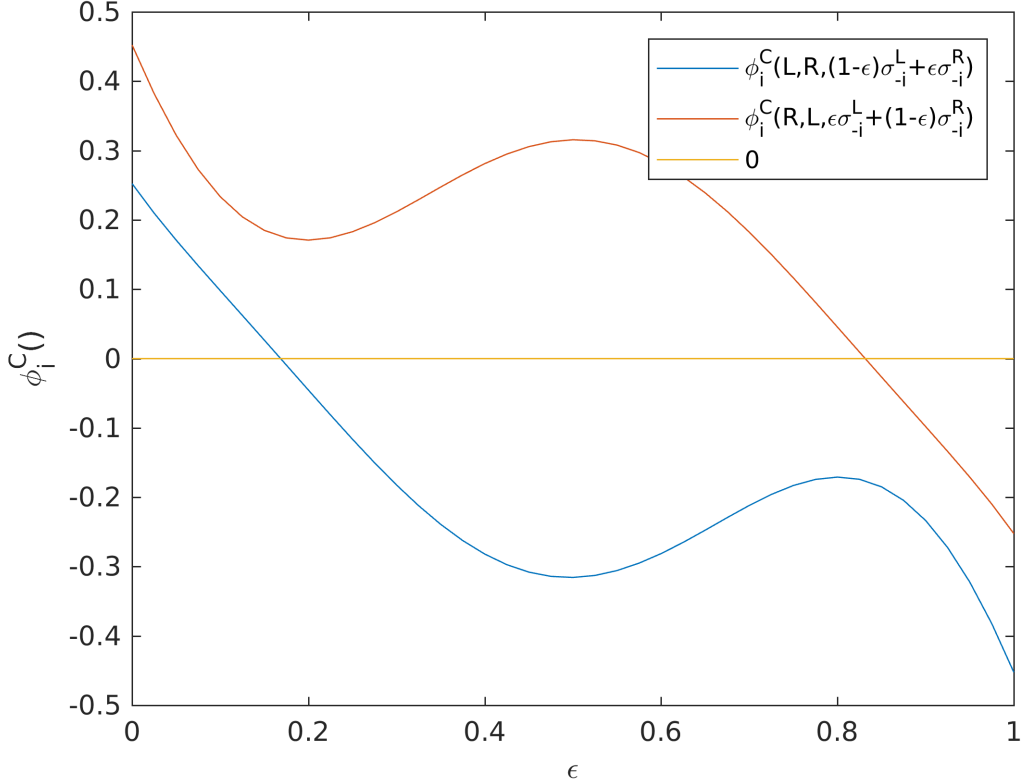


Figure 1: Plot of $\phi(\cdot)$ for $(n, e, K_1, K_2, K_3) = (21, 3, -1, -0.5, 0.3)$

We now show that for some $i \in \Omega^R$, or said differently, for a selection of parameter values of n , e , K_1 , K_2 and K_3 , σ^R is not immune to uncorrelated errors at level $\frac{1}{2}$ w.r.t. σ^L . To illustrate these examples, and also coincidentally add intuition about our ϕ_C^i function in the case of uncorrelated errors, Figures 1 to 6 show the graphs of $f_i(\varepsilon) = \phi_i^C(\sigma_i^L, \sigma_i^R, (1 - \varepsilon)\sigma_{-i}^L + \varepsilon\sigma_{-i}^R)$ and $-f_i(1 - \varepsilon) = \phi_i^C(\sigma_i^R, \sigma_i^L, (1 - \varepsilon)\sigma_{-i}^R + \varepsilon\sigma_{-i}^L)$ for different choices of the (model dependent) vector of parameters (n, e, K_1, K_2, K_3) , while θ is fixed to $\theta = 1$. The values of K_1, K_2, K_3 chosen in the last three figures were calculated based on an explicit u function. Namely, we chose $u(x) = -\frac{1}{2}x^2$, $t_L = 0.2$, $t_R = 0.8$, $t_{LE}^s = 0.4$, $t_{RE}^s = 0.6$, $t_i = 0.65$, and $\bar{d} = 0.1$. All plots were generated using MATLAB.

We see that for the three cases where the K constants were chosen according to a specific utility function and policy values, the first root of $\phi_i^C(\sigma_i^R, \sigma_i^L, (1 - \varepsilon)\sigma_{-i}^R + \varepsilon\sigma_{-i}^L)$

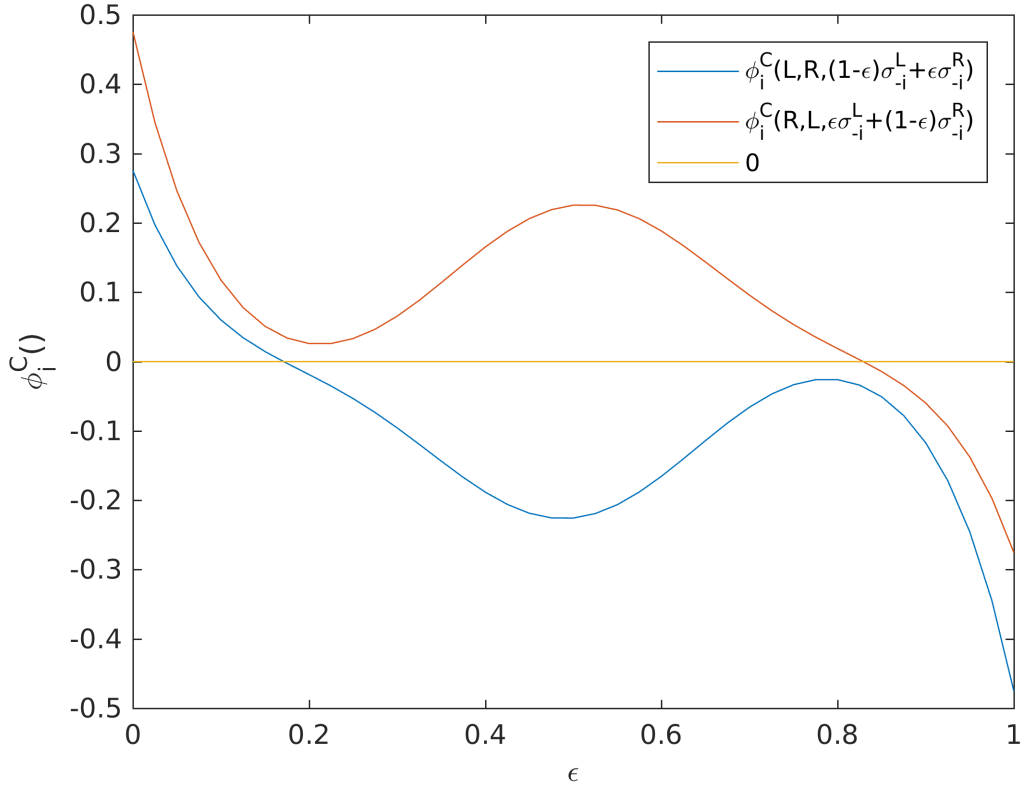


Figure 2: Plot of $\phi(\cdot)$ for $(n, e, K_1, K_2, K_3) = (41, 6, -1, -0.5, 0.3)$

occurs before 0.2. That is, σ^R is immune to uncorrelated errors w.r.t. σ^L at most at a level smaller than 0.2. Moreover, in the case where $(K_1, K_2, K_3) = (-1, -0.5, 0.3)$, when n and e are large enough $((n, e) = (81, 12))$, we also see that σ^R can be immune to uncorrelated errors w.r.t. σ^L at most at a level smaller than 0.2.

C.5 Proof of Proposition 6

We start showing non-comparability of strategic coordinated voting equilibria with respect to 'domination according to maximum entropy in equilibria' and 'domination in the presence of correlated errors'.

As before, by Lemma 3, 'domination according to maximum entropy in equilibria' and 'domination in the presence of correlated errors' are equivalent. Thus, it is sufficient to show that

$$\phi_i^C(\sigma^L, \sigma^R, \sigma_{-i}^L) - \phi_i^C(\sigma^R, \sigma^L, \sigma_{-i}^R)$$

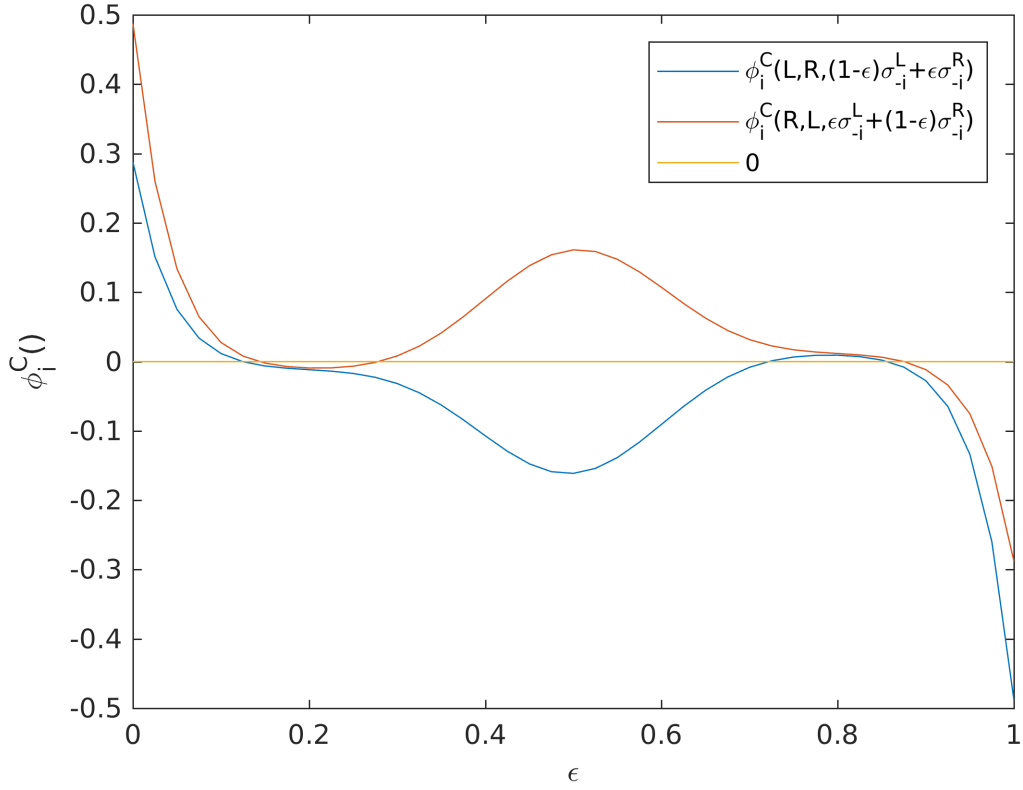


Figure 3: Plot of $\phi(\cdot)$ for $(n, e, K_1, K_2, K_3) = (81, 12, -1, -0.5, 0.3)$

is > 0 for $i \in \Omega^L$ and is < 0 for $i \in \Omega^R$. We have:

$$\begin{aligned}\phi_i^C(\sigma^L, \sigma^R, \sigma_{-i}^L) &= u(|t_i - t_L|) - \theta \left(\frac{1}{2} + \frac{1}{2n} \right) - u(|t_{LE}^s - t_i|) + \theta \left(\frac{1}{2} - \frac{1}{2n} \right) + \bar{d}, \\ \phi_i^C(\sigma^R, \sigma^L, \sigma_{-i}^R) &= u(|t_i - t_R|) - \theta \left(\frac{1}{2} + \frac{1}{2n} \right) - u(|t_{RE}^s - t_i|) + \bar{d} + \theta \left(\frac{1}{2} - \frac{1}{2n} \right).\end{aligned}$$

Thus

$$\begin{aligned}\phi_i^C(\sigma^L, \sigma^R, \sigma_{-i}^L) - \phi_i^C(\sigma^R, \sigma^L, \sigma_{-i}^R) \\ = u(|t_i - t_L|) - u(|t_i - t_{LE}^s|) + u(|t_i - t_{RE}^s|) - u(|t_i - t_R|).\end{aligned}$$

By the usual method of separating the domain of t_i , one checks that the above is strictly decreasing in t_i . Then, note that for $t_i = t_m$ the expression is 0 by symmetry. Thus, we must have > 0 for $t_i < t_m$ and < 0 for $t_i > t_m$.

Next we prove non-comparability of strategic coordinated voting equilibria with respect to 'domination in the presence of uncorrelated errors'.

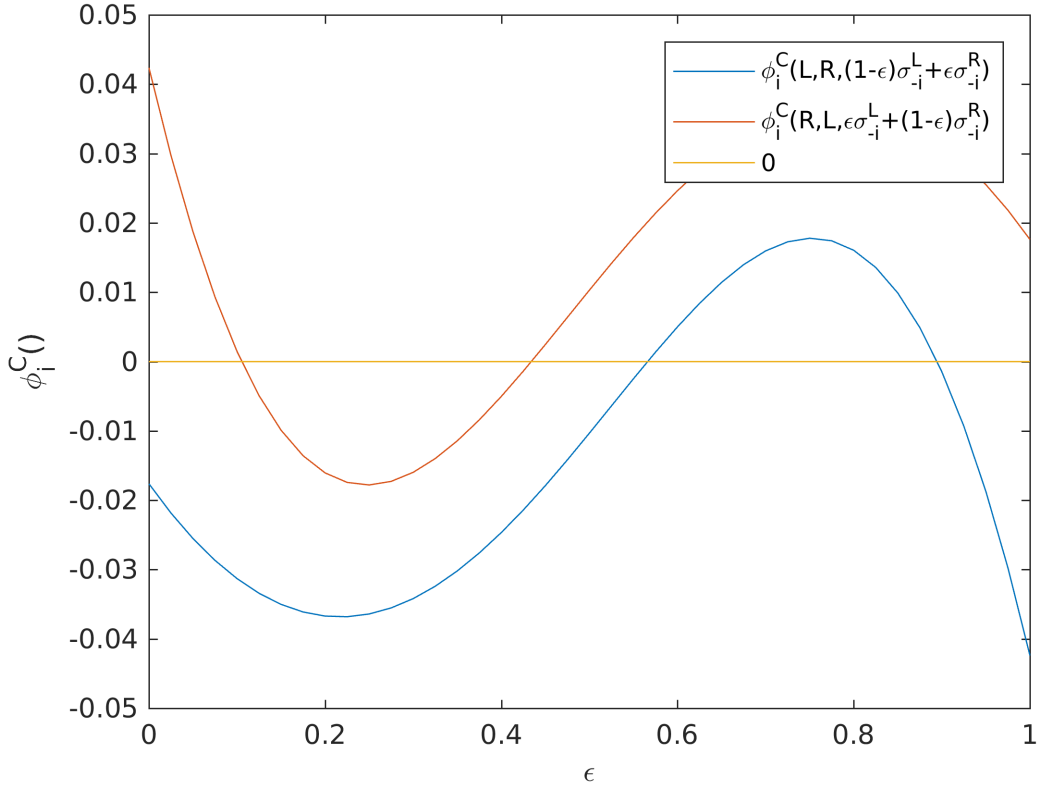


Figure 4: Plot of $\phi(\cdot)$ for $(n, e, K_1, K_2, K_3) = (21, 3, -0.03, -0.09, 0.03)$

Lemma 4 already established that $\sigma^R \succ_i^\perp \sigma^L \forall i \in \Omega^R$. We take the results derived in Lemma 4:

$$\begin{aligned} f_i(\varepsilon) &:= \phi_i^C(\sigma_i^L, \sigma_i^R, (1-\varepsilon)\sigma_{-i}^L + \varepsilon\sigma_{-i}^R) \\ &= 2p(B_{1-\varepsilon}^{2e} = e)K_1 + \{\varepsilon^{2e} + (1-\varepsilon)^{2e}\} [K_2 + K_3], \end{aligned}$$

where recall that

$$\begin{aligned} K_1 &:= u(|t_i - t_{LE}^s|) - u(|t_{RE}^s - t_i|), \\ K_2 &:= u(|t_{RE}^s - t_i|) - u(|t_R - t_i|) - \bar{d}, \\ K_3 &:= u(|t_i - t_L|) - u(|t_i - t_{LE}^s|) + \bar{d}, \end{aligned}$$

and show that $f_i(\varepsilon) + f_i(1-\varepsilon) > 0 \forall i \in \Omega^L$, which by Proposition 4 (iii), proves that $\sigma^L \succ_i^\perp \sigma^R \forall i \in \Omega^L$. For $i \in \Omega^L$, we have $K_1 > 0$, $K_2 < 0$ and $K_3 > 0$. In a completely analogous manner to the way we showed $K_2 + K_3 < 0$ if $i \in \Omega^R$, we can show that $K_2 + K_3 > 0$ if $i \in \Omega^L$. This then completes the proof as all other quantities in the function above are greater than 0.

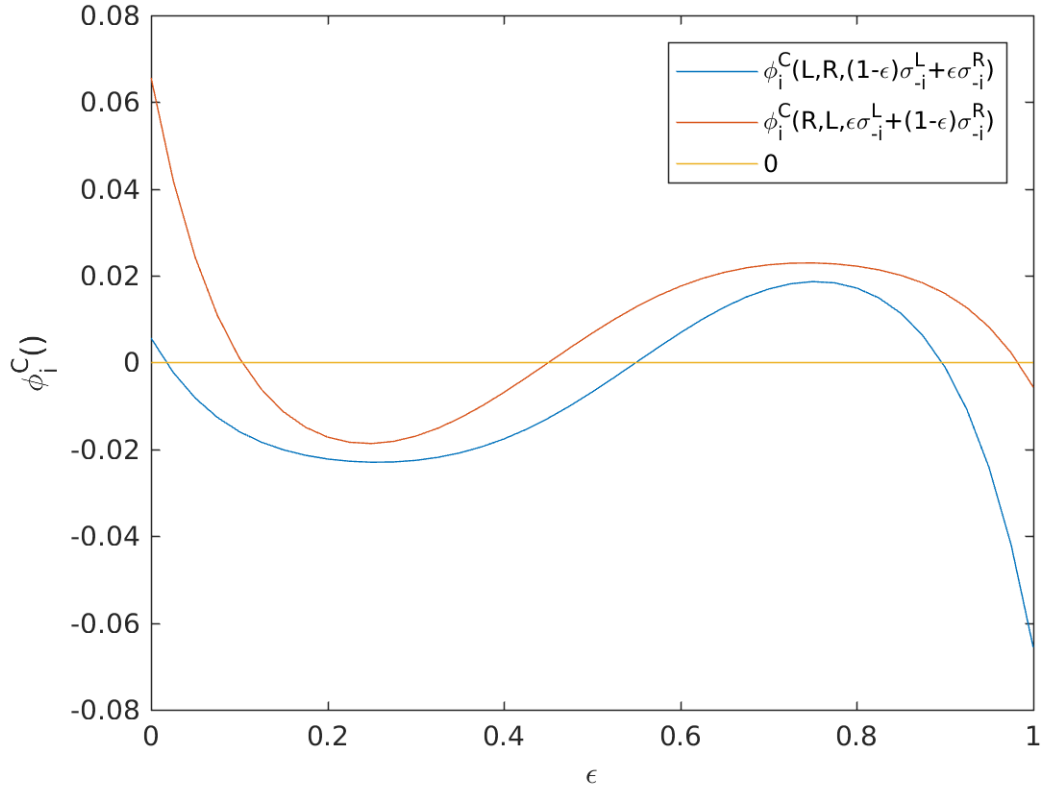


Figure 5: Plot of $\phi(\cdot)$ for $(n, e, K_1, K_2, K_3) = (41, 6, -0.03, -0.09, 0.03)$

Finally we show non-comparability of strategic coordinated voting equilibria with respect to 'domination according to maximum entropy in equilibria'.

By definition, it is sufficient to show that

$$\phi_i^C \left(\sigma^L, \sigma^R, \frac{1}{2}\sigma_{-i}^L + \frac{1}{2}\sigma_{-i}^R \right)$$

is > 0 for $i \in \Omega^L$ and < 0 for $i \in \Omega^R$.

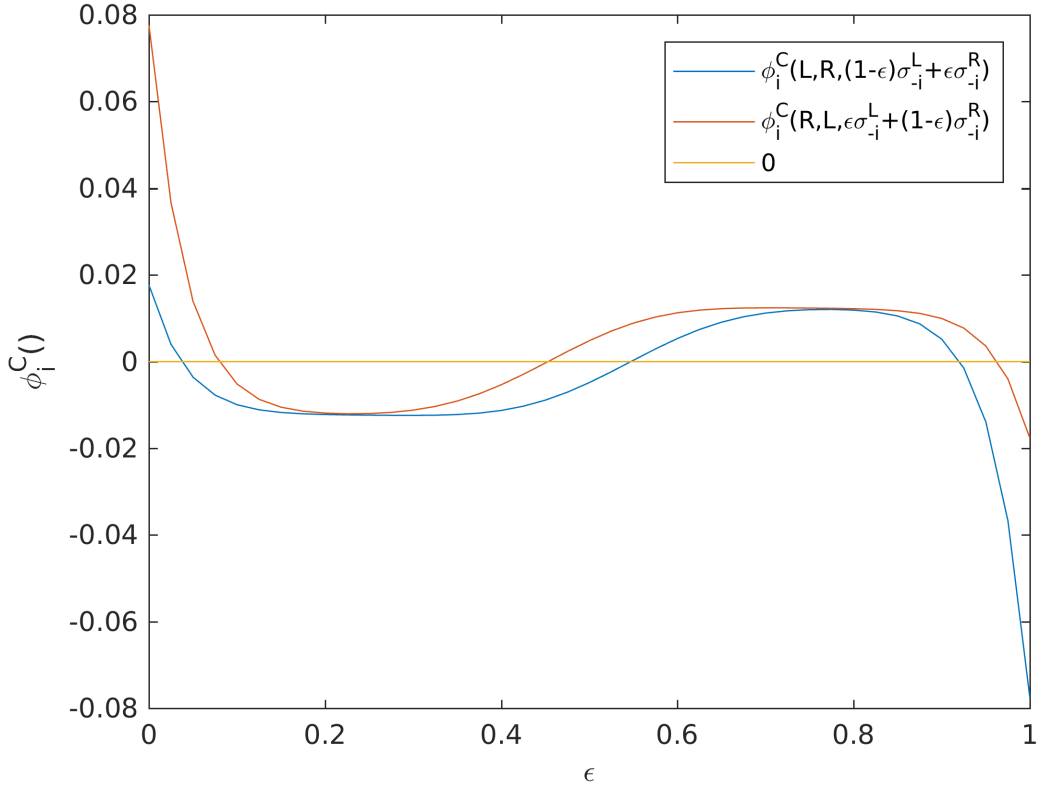


Figure 6: Plot of $\phi(\cdot)$ for $(n, e, K_1, K_2, K_3) = (81, 12, -0.03, -0.09, 0.03)$

We have $\chi_{-i}^{\frac{1}{2}\sigma^L + \frac{1}{2}\sigma^R} \sim \text{Bin}(\frac{1}{2}, 2e)$. Thus:

$$\begin{aligned}
\phi_i^C\left(\sigma^L, \sigma^R, \frac{1}{2}\sigma_{-i}^L + \frac{1}{2}\sigma_{-i}^R\right) &= \\
\sum_{j=1}^{e-1} p(\chi_{-i}^{\frac{1}{2}\sigma^L + \frac{1}{2}\sigma^R} = j) &\left[u(|t_{RE}^s - t_i|) - \bar{d} - \theta\left(\frac{1}{2} + \frac{1}{2n} - \frac{j}{n} - \frac{1}{n}\right) \right. \\
&\quad \left. - u(|t_{RE}^s - t_i|) + \bar{d} + \theta\left(\frac{1}{2} + \frac{1}{2n} - \frac{j}{n} + \frac{1}{n}\right) \right] \\
+ \sum_{j=e+1}^{2e-1} p(\chi_{-i}^{\frac{1}{2}\sigma^L + \frac{1}{2}\sigma^R} = j) &\left[-2\frac{\theta}{n} \right] \\
+ p(\chi_{-i}^{\frac{1}{2}\sigma^L + \frac{1}{2}\sigma^R} = 0) &\left[u(|t_{RE}^s - t_i|) - \bar{d} - \theta\left(\frac{1}{2} - \frac{1}{2n}\right) - u(|t_R - t_i|) + \theta\left(\frac{1}{2} + \frac{1}{2n}\right) \right] \\
+ p(\chi_{-i}^{\frac{1}{2}\sigma^L + \frac{1}{2}\sigma^R} = e) &\left[u(|t_{LE}^s - t_i|) \right. \\
&\quad \left. - \bar{d} - \theta\left(\frac{1}{2} + \frac{1}{2n}\right) - u(|t_{RE}^s - t_i|) + \bar{d} + \theta\left(\frac{1}{2} + \frac{1}{2n}\right) \right]
\end{aligned}$$

$$\begin{aligned}
& + p(\chi_{-i}^{\frac{1}{2}\sigma^L + \frac{1}{2}\sigma^R} = 2e) \left[u(|t_L - t_i|) - \theta \left(\frac{1}{2} + \frac{1}{2n} \right) - u(|t_{LE}^s - t_i|) + \bar{d} + \theta \left(\frac{1}{2} - \frac{1}{2n} \right) \right] \\
& = \sum_{j=1}^{e-1} p(\chi_{-i}^{\frac{1}{2}\sigma^L + \frac{1}{2}\sigma^R} = j) \left[+2\frac{\theta}{n} \right] + \sum_{j=e+1}^{2e-1} p(\chi_{-i} = j) \left[-2\frac{\theta}{n} \right] \\
& + \left(\frac{1}{2} \right)^{2e} \left[u(|t_L - t_i|) - u(|t_{LE}^s - t_i|) + u(|t_{RE}^s - t_i|) - u(|t_R - t_i|) \right] \\
& + p(\chi_{-i}^{\frac{1}{2}\sigma^L + \frac{1}{2}\sigma^R} = e) \left[u(|t_{LE}^s - t_i|) - u(|t_{RE}^s - t_i|) \right] \\
& = \left(\frac{1}{2} \right)^{2e} [K_2 + K_3] + p(\chi_{-i}^{\frac{1}{2}\sigma^L + \frac{1}{2}\sigma^R} = e)K_1,
\end{aligned}$$

where in the last step we used the fact that the binomial distribution is exactly symmetric about its mean, when the number of trials is even (i.e. there are $2e + 1$ possible values $\chi_{-i}^{\frac{1}{2}\sigma^L + \frac{1}{2}\sigma^R}$ can take). As mentioned in the proof of (vi) above, $K_1 > 0$ and $K_2 + K_3 > 0$ if and only if $t_i < \frac{1}{2}$ i.e. $i \in \Omega^L$ and this gives us the desired result.