How Does or Can Democracy Cope with Extreme Views?*

Hans Gersbach  
CER-ETH  
Center of Economic Research at ETH Zurich and CEPR  
8092 Zurich, Switzerland  
hgersbach@ethz.ch

Vincent Lohmann  
CER-ETH  
Center of Economic Research at ETH Zurich  
8092 Zurich, Switzerland  
lohmannv@ethz.ch

Maik T. Schneider  
Department of Economics  
at University of Bath  
Claverton Down, BA2 7AY Bath, UK  
m.t.schneider@bath.ac.uk

Oriol Tejada  
CER-ETH  
Center of Economic Research at ETH Zurich  
8092 Zurich, Switzerland  
tejada@mip.mtec.ethz.ch

This Version: May 27, 2019

Abstract

We investigate how democracy can cope with extreme views held by a minority of citizens. If parties can credibly preclude a coalition with an extremist party after an election, this might prevent extreme policy shifts and increase social welfare. We establish general conditions under which such ‘Coalition-Preclusion Promises’ (CPPs) lead to welfare improvements. We find that while CPPs carry welfare improvements under relatively weak conditions on the extent of the policy change envisioned by the extremist party, these welfare gains can only be realized if parties are able and have sufficient incentives to use CPPs. This depends crucially on the parties’ beliefs about voters’ coordinating their votes on the party making a CPP. The coordination problem among strategic voters exhibits multiple equilibria. We extend typical equilibrium selection criteria to apply them to this voting game. We find that CPPs can be an effective tool to prevent extremist parties from entering government.

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

JEL: D72, D82, H55

*We would like to thank Carlos Alos-Ferrer for helpful comments. All errors remain our own.
1 Introduction

In many developed democracies, traditionally, 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.\(^1\) 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.\(^2\) 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 extreme cultural policies, when placed within the UK political spectrum,\(^3\) such as not recognizing same-sex marriage and considerably reducing abortion rights.\(^4\) 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

---

\(^1\)Consider for example Spain, where since the reinstatement 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.

\(^2\)Most recently 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).

\(^3\)As 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.

\(^4\)Interestingly, 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.
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 Catalunia, 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.

This motivates our investigation of whether and how “Coalition-Preclusion Promises” can prevent extreme policy shifts due to extremist parties entering government. By making a “Coalition-Preclusion Promise” (CPP) conventional parties can credibly preclude entering a coalition government with a (usually extremist) party after the election.

We formulate a general two-stage extensive form game encapsulating the central characteristics of a political process. 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 minority of citizens supports the extreme party. Parties, i.e. their leading politicians, 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 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 making 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 gains 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 announcing a CPP. 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 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 not. For simplicity, and because our focus in this paper is on the political theory literature, we consider only 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 show that under very mild conditions, the first criterion is equivalent to the third, and the second to the fourth.

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 CPPs 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 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 tend to 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

---

5Throughout 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.
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 10 we discuss how such CPPs may emerge or could be introduced.

The paper is organised as follows. First, we discuss the relation of our paper to the literature. Then we introduce the comprehensive political game where conventional parties can or cannot credibly exclude the extremist party from a coalition government. In Section 4, we extract the essence of the game in a reduced ‘core game’ and characterize all subgame perfect Nash equilibria in Section 5. (to be written)

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. 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

\(^6\)A discussion of the literature on binding commitments in election campaigns can be found in Gersbach et al. (2019).
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 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.\footnote{More 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.}

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.

\section{The Model}

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

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

Second, we denote the finite set of citizens by \( \mathcal{N} \), with cardinality \( n = 2m+1 \).\footnote{We assume that \( n \) is odd number, which eases the presentation considerably.} W.l.o.g, we let \( \mathcal{N} = \{0, 1, \ldots, 2m\} \). Each citizen \( i \in \mathcal{N} \) is characterized by his utility over the policy space. Besides policies, voters also care about perks in office. Specifically, we assume that utilities are additively separable in the two policy dimensions. That is for
each $i \in \mathcal{N}$,

$$U_i(t, d, s_G) := u(|t - t_i|) - \delta_i d - \theta s_G,$$

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 policy by $t_m$. Without loss of generality we assume $t_m = \frac{1}{2}$.\textsuperscript{9} 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, $\theta s_G$, where $\theta > 0$ and $s_G$ is a measure of the number of perk-seeking politicians defined in detail on page 10, captures the disutility associated with perks obtained by members in the government. We assume that for any given two-dimensional policy, voters prefer governments with smaller support rather than with larger support, since large governments use more resources. Of course, if a larger government implements a different policy than a smaller one, a majority of voters might prefer the larger government.\textsuperscript{10}

Third, there are three parties, denoted by $L$, $R$ and $E$. Each party’s political orientation is reflected by its most preferred policy $p \in \mathcal{P}$, which we also refer to as the party platform. On the one hand, we assume that $p_L = (t_L, 0)$ and $p_R = (t_R, 0)$ are the platforms of parties $L$ and $R$ respectively, which are henceforth referred as conventional parties. We assume that $t_L < t_m = \frac{1}{2} < t_R$, so party $L$ (resp. $R$) is the left-wing (resp. right-wing) party, and $t_R + t_L = 1$, so conventional parties have symmetric positions in the $\mathcal{T}$-policy dimension around the median $t_m = \frac{1}{2}$. On the other hand, we consider that $p_E = (t_E, \bar{d})$ is the platform of party $E$, which is henceforth referred as the extreme party. Let the indicator $I_j$ be defined as follows:

$$I_j = \begin{cases} 
1 & \text{if party } j \text{ takes part in the government and } j \in \{L, R\}, \\
0 & \text{otherwise.}
\end{cases}$$

Then, the utility of a party $j \in \{L, R, E\}$ with platform $p_j = (t_j, d_j)$ is given by

$$V_j \left( t, d, I_j \cdot \frac{s_j}{s_G} \right) = u(|t - t_j|) - \alpha_j d + I_j \cdot 2\tau_j \frac{s_j}{s_G},$$

where

$$\alpha_j = \begin{cases} 
+1 & \text{if party } j \in \{L, R\}, \\
-1 & \text{otherwise,}
\end{cases}$$

\textsuperscript{9}If 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 $\mathcal{T} = [0, 1]$.

\textsuperscript{10}That is, we capture the issue of “stability” of a government via the policy space $\mathcal{T}$. 
$s_j$ is the share of parliamentary seats for party $j$, and $s_G$ the share of parliamentary seats of the government coalition which are relevant for perks. The first two parts of $V_j$ capture the utility a representative voter with the ideal point of party $j$ receives. For each party $j$, the parameter $\tau_j$ captures the (exogenous) ability of party $j$ to extract perks when being in government according to its relative share in the government. Unlike voters, political parties do like perks. We denote by $G \subseteq \{L, R, E\}$ the set of parties that take part in the government. We assume that only conventional parties are able to extract perks, e.g. because they are better connected to powerful groups or because the unique leitmotiv of party $E$ is to implement the extreme policy. As a consequence, the share $s_G$ of the coalition in the parliament that gives support to the government and is relevant for perks can be written as

$$s_G = \sum_{j \in G \setminus \{E\}} s_j.$$  \hspace{1cm} (3)

The misalignment between the interests of the majority of voters (represented by the median voter) and the interests of the two conventional parties is fueled in our model by three different components. Indeed, for $j \in \{L, R\}$,

(i) party $j$’s platform is $t_j$, which is different from $t_m$.

(ii) party $j$ is able to extract perks, but perks associated with the ruling coalition generate disutility for all voters.

(iii) the parameter $\tau_j$ is sufficiently large to ensure that party $j$ prefers forming a government without the other conventional party in the circumstances analysed below, even when the grand coalition is the most preferred government coalition for a majority of voters.

4 The Political Game

The political game involves three main stages:

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 may simply be cheap talk and thus may have no impact on subsequent elections and government formation. They may also have commitment power and we discuss in Section 10 how the commitment power of such CPPs can be strengthened by institutional means. At this stage, we simply allow both polar cases—when CPPs have either no, or full commitment power.

We will discuss in which circumstances democracy can cope better with extreme views. We focus on the circumstances in which the two conventional parties might promise to exclude the extreme party from government. We assume that the extreme party is willing to enter government and thus to form a government with a conventional party. Moreover, we assume that a conventional party excluding the other conventional party is always considered cheap talk. The CPP of party \( j \in \{L, R\} \) is denoted by \( C_j \in \{\emptyset, E\} \).\(^{11}\) 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 \( x_i \in \{L, R, E\} \).\(^{12}\) 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.\(^{13}\) To obtain seats in parliament, parties need a certain share \( s_{\text{E}}^* \) of total votes.\(^{14}\)

In the third stage, the government formation process occurs as follows: If a conventional party obtains a majority of seats, it will form a single-party government, which in turn will select a policy that will be approved in parliament by a vote of confidence. Otherwise, both conventional parties will try to establish a coalition government in line with the following stages:

\(^{11}\)We abuse notation, and always write \( C = E \), instead of \( C = \{E\} \).

\(^{12}\)For simplicity, we do not consider abstention.

\(^{13}\)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” \( T : |N|^3 \rightarrow SP^3 \), 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.

\(^{14}\)We assume that if \( s_{\text{E}}^* = 0 \) then a share \( s > s_{\text{E}}^* = 0 \) is required to enter parliament, while if \( s_{\text{E}}^* > 0 \), then a share \( s \geq s_{\text{E}}^* \) is required.
Stage 3.1: Proposal Round. Conventional parties that have not excluded $E$ in their contract simultaneously offer a coalition government to $E$ by suggesting a policy $p = (t, d)$ for implementation.

Stage 3.2: Acceptance Round. Party $E$ decides which offer to accept (if any) from the policies proposed.

Stage 3.3: Grand Coalition Bargaining. If both conventional parties have excluded $E$ in Stage 1, $E$ rejects all offers or no proposal is made to $E$, then a coalition between the two conventional parties will occur.

Stage 3.4: Vote of Confidence. The proposed coalition gains power to execute some agreed policy $p = (t, d)$ if it receives a majority of votes in parliament.

Stage 3.5: Caretaker Government. If the vote of confidence fails, then a “caretaker government” will take over the duties of the executive branch.

Typically, the caretaker government will consist of bureaucrats ensuring that operations in the executive branch keep running. This means that a caretaker government would stick to $d = 0$ and implement some policy $t \in T$. Nevertheless, we assume that parties and voters suffer a very large utility loss when a caretaker government runs the executive branch. Accordingly, both conventional parties are always better off forming a grand coalition, which is in turn preferred to a caretaker government by the voters. Moreover, when no party has a majority, any two parties that reach an agreement on some policy $p$ at stages 3.2 or 3.3 will see their agreed policy receive the majority in the vote of confidence. These two observations permit us simplifying the government formation subgame, which we henceforth consider to be made up only of Stages 3.1, 3.2, and 3.3.

4.1 Further assumptions

To facilitate the analysis of the model, we make a number of further assumptions.

4.1.1 On the information uncertainty and the size of the electorate

Unless stated to the contrary, we henceforth impose the following assumptions regarding the uncertainty associated with the support for the extreme party: First, we assume
for simplicity that the extreme party will always obtain a certain non-stochastic number of votes in the election, say $2e$, and enter parliament. Thus we have $s_E = \frac{2e}{2m+1}$. In this paper we focus on cases where an extreme minority party exists and in particular enters parliament. As such, either $0 \leq s^*_E < s_E < \frac{1}{2}$ or $0 < s^*_E \leq s_E < \frac{1}{2}$ is assumed. This immediately implies that $2e + 2e + 1 = 4e + 1 \leq n$. Because $t_L + t_R = 1$ and the median voter $m$ has $t_m = \frac{1}{2}$, the composition of the electorate is completely characterised 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 $e = 2$.

Second, we assume that the ideal points of the $2e$ extreme voters regarding the policy in $T$ are symmetrically distributed around the median $t_m = \frac{1}{2}$. That is the representative voter of the extreme party would like to implement $d = \bar{d}$ and $t = t_m = \frac{1}{2}$. In turn, because the total number of voters is odd and no two voters are identical, this implies that the median of the conventional voters is still uniquely attained and equals $m$, or in other words the median voter of all voters is a conventional voter, and it is also the median voter when considering only the subset of all conventional voters.

Note that both of these assumptions would be satisfied in the limit as $n \to \infty$ if we let voters’ types in the $T$-dimension be i.i.d. drawn from a distribution on $[0, 1]$ which is symmetric about $\frac{1}{2}$, and their types in the $D$-dimension be i.i.d. drawn (independently of the draws in the $T$-dimension) from a Bernoulli distribution with success probability $p = s_E$ (where we identify success with being an extreme voter with $\delta_i = -1$).

### 4.1.2 On voters’ behavior

We impose some assumptions that constrain the behavior of voters. On the one hand, we assume that extreme voters always vote for $E$ no matter what. This can be justified by ideological considerations, see Benabou (2008), as the extreme party is the only one which advocates for a very specific policy change.

On the other hand, unlike extreme voters, conventional voters can either vote sincerely for the conventional party whose platform is closer to their preferred policy, or vote strategically otherwise for the other conventional party. Besides the assumption that conventional voters never vote for the extreme party, the behavior of these voters is, however, not completely unrestricted. More specifically, we impose a series of as-
sumptions. First, we assume that conventional voters derive a certain small, non-zero disutility when voting strategically. Second, we consider that such disutility increases as voters’ preferred policy gets farther away from the median policy in the $T$-policy. Third, we assume that if the same voting outcome can be attained by two different voting profiles, only the voting profile which minimizes the costs associated with strategic voting will actually emerge.

Combining the three assumptions with the property $|t_m - t_L| = |t_m - t_R|$ implies the following: only sets of minimal size composed of conventional voters may vote strategically. More specifically, these coalitions are either empty or they include exactly $e$ voters. Note that our assumptions imply that except for $2e + 1$ conventional voters, all voters vote sincerely. The set of citizens potentially voting strategically consists of $e$ voters to the left and $e$ voters to the right of the median voter, and the median voter himself.

Finally, we assume as a tie-breaking rule that if the median voter is indifferent between voting for $R$ and $L$, he votes for either party with probability $\frac{1}{2}$.

4.1.3 On parties’ behavior

We also impose some assumptions that constrain the behavior of parties. On the one hand, we assume $d = \bar{d}$ acts as an ideological constraint on the mobility of the extreme party (see Mueller (2003) and Benabou (2008)). This translates into the assumption that the extreme party will not make any compromise on the $D$-policy dimension if it enters the government. Finally, we assume as a tie-breaking rule that in case of indifference, a conventional party excludes $E$ in its CPP.

4.1.4 On the government formation process

Beyond the mobility constraint that defines the extreme party, we consider several assumptions regarding the government formation process. 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 $k \in \{L, R\}$, will implement policy $p^*_kE = (t^*_kE, \bar{d})$. Moreover, we assume that 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 $E$.\footnote{Possible justifications for this assumption are, that the party with more seats can always offer an \( \varepsilon \) 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 $E$.}

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 $k \in \{L, R\}$ is able to form a coalition with $E$, the extreme party has weak bargaining power, and the resulting coalition will implement policy $p^{w}_{kE} = (t^{w}_{kE}, \bar{d})$.

We assume that
\[
t_L \leq t^{w}_{LE} \leq t^s_{LE} \leq t_m \leq t^s_{RE} \leq t^{w}_{RE} \leq t_R
\]
and
\[
t^s_{LE} + t^s_{RE} = t^{w}_{LE} + t^{w}_{RE} = 1.
\]

Third, when both conventional parties bargain over the desired policy $p = (t, d)$ for implementation, they do so by maximizing
\[
s_L \cdot V_L \left( t, d, \frac{s_L}{s_L + s_R} \right) + s_R \cdot V_R \left( t, d, \frac{s_R}{s_L + s_R} \right). \footnote{This bargaining procedure yields the Nash Bargaining solution with bargaining power proportional to the share in the parliament.}
\]

The above assumptions lead to the following result.

**Proposition 1**

In any equilibrium of the political game $G$ defined above, 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, i.e. 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 $p^s_{jE}$ 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 $p^{w}_{jE}$ 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 policy maximizing (6), which we henceforth denote by \(t(s_L, s_R)\), will be implemented. More precisely, since both parties’ utility can be increased by moving \(t\) to a more central position if \(t \notin [t_L, t_R]\), we have,

\[
t(s_L, s_R) = \arg\max_{t \in [t_L, t_R]} s_L u(t - t_L) + s_R u(t_R - t)
\]

(which is always in \([t_L, t_R]\)). In particular, if \(u(x) = -\frac{1}{2}x^2\), for \(x \geq 0\), then \(t(s_L, s_R)\) is just a convex combination of \(t_L\) and \(t_R\):

\[
t(s_L, s_R) = \frac{s_L}{s_L + s_R}t_L + \frac{s_R}{s_L + s_R}t_R.
\]

In this case, \(s_G = 1 - s_E\).

4.2 The “Core Game”

From the above considerations, it will suffice to analyze the dynamic game denoted by \(G^-E\) and described as follows: The set of players comprises the conventional parties \(R\) and \(L\), on the one hand, and the median voter, \(m\), plus the \(e\) conventional voters which are closest to \(m\) on both sides of the spectrum. For the sake of notation, we rename the set of the considered voters to \(\Omega = \{0, 1, \ldots, 2e\}\), and we let \(\Omega^L = \{i \in \Omega : t_i < t_m\}\) and \(\Omega^R = \{i \in \Omega : t_i > t_m\}\) denote the corresponding sets of left-wing voters and right-wing voters. On the other hand, if CPPs are available, each party’s strategy set is \(\emptyset, E\), while each voter’s strategy set is \(\{L, R\}\). The political game involves two stages:

**Stage 1**: Potential exclusion of extreme party from government.

**Stage 2**: Election.

In Stage 1 of \(G^-E\), both conventional parties, \(L\) and \(R\), simultaneously choose from \(\emptyset, E\). In Stage 2 of \(G^-E\), all voters from \(\Omega\) simultaneously cast a vote for either \(L\) or \(R\).\(^{17}\) We denote a strategy profile of the CPP stage by \(C = (C_L, C_R) \in \emptyset \times \emptyset\) and at the voting stage by \(\sigma = (\sigma_i)_{i \in \Omega} \in \{L, R\}^\Omega\). After Stage 2, payoffs are derived from (1) and (2) according to Proposition 1. An equilibrium in our model is any

\(^{17}\)To focus on the interaction between parties’ CPP and voters’ decisions, we assume that no abstention occurs.
subgame perfect Nash equilibrium of $G^E$. For each voter $i \in \Omega$, we let $\sigma_i = pLR$ denote the mixed strategy of $i$ according to which he votes $L$ with probability $p$ and $R$ with probability $1 - p$.

Let henceforth denote $t^\Omega_L = \min_{i \in \Omega} t_i$ and $t^\Omega_R = \max_{i \in \Omega} t_i$. We assume

$$t^\Omega_L + t^\Omega_R = 1. \quad (8)$$

Then $\eta = t^\Omega_R - t_R = t_L - t^\Omega_L$ relates the preferred policies in the $T$-policy dimension of the conventional voters who are most different from the median voter to the preferences of the conventional parties. Note that $\eta$ can be either positive or negative.

5 Solving the “Core Game”

In this section, we solve the game $G^E$ by backward induction. Hence, 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.

5.1 The solution to Stage 2

According to the assumptions in Section 4.1, the possibilities for a government are: (a) a conventional single-party Government, (b) a two-party coalition government containing the extreme party, and (c) a grand-coalition government. Moreover, it trivially follows from these assumptions that if $(C, \sigma)$ is an equilibrium of $G^E$, then $\sigma$ has to be of the following types:

- Case 1: $\sigma = \sigma^L := (L, \ldots, L)$.
- Case 2: $\sigma = \sigma^{LR} := (L, \ldots, L, eLR, R, \ldots, R)$.
- Case 3: $\sigma = \sigma^R := (R, \ldots, R)$.

The support for each of the conventional parties is depicted in Table 1. We 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.
Next, we investigate which of the above voting profiles may arise in equilibrium of the voting subgame depending on the CPPs made in Stage 1. Accordingly, we consider four different scenarios. For each \( i \in \Omega \), we let \( EU_i(\sigma) \) be the (expected) utility obtained by \( i \) when voters act according to the voting profile \( \sigma \). Voters are risk-neutral. As usual, for each \( i \in \Omega \) and voting profile \( \sigma, \sigma_{-i} \in \{L, R\}^{\Omega \setminus \{i\}} \) denotes the strategy profile of all players but \( i \). Lastly, let \( \mathcal{E}^C \) denote the set of equilibria in the subgame of \( G^{-E} \) that starts at Stage 2, for a given choice of CPP, say \( C \).

For notational convenience, we introduce the following parameters. First, we define the party polarization of the voting game as the difference between the preferred platforms of the two parties, i.e.,

\[
\pi := t_R - t_L.
\]  

(9)

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 \( T \)-dimension when bargaining with \( E \) in the case where the latter has strong bargaining power w.r.t. to their platforms, i.e.,

\[
\beta^s := t_R - t_{RE}^s = t_{LE}^s - t_L.
\]  

(10)

When \( \beta^s \) is large, the conventional parties are not able to shift the policy in the \( T \)-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 \( T \)-dimension when bargaining with \( E \) in the case where the latter has weak bargaining power w.r.t. to their platforms, i.e.,

\[
\beta^w := t_R - t_{RE}^w = t_{LE}^w - t_L.
\]  

(11)

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

\[ \gamma := \frac{1}{2} - s_E. \]  

(12)

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. In particular, we assume the following:

**Polarization Conditions (PCS)**

\begin{align*}
(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).
\end{align*}

The conditions (PCS) are used throughout the paper, because they 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 \( \bar{d} \).\(^{18}\) 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.

**Corollary 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 \( i \in \Omega \):

\begin{align*}
(i) & \quad \bar{d} \geq s_E \theta \pm u(|t^j_{LE} - t_i|) - u(|t_R - t_i|). \\
(ii) & \quad \bar{d} \geq s_E \theta \pm u(|t^j_{LE} - t_i|) - u(|t_L - t_i|). \\
(iii) & \quad \bar{d} \geq s_E \theta \pm u(|t^j_{RE} - t_i|) - u(|t_R - t_i|). \\
(iv) & \quad \bar{d} \geq s_E \theta \pm u(|t^j_{RE} - t_i|) - u(|t_L - t_i|).
\end{align*}

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

\(^{18}\)Note that if we assume that \( \beta^w + \eta > 0 \), that is, that \( t^w_L < t^w_{LE} \) and conversely \( t^w_R < t^w_{RE} \), then the Weak Polarization Condition implies the Strong Condition since \( u \) is assumed to be decreasing.
Moreover, we consider the following upper-bound for the sensitivity of voters to perks:

**Sensitivity-to-Perks Condition (SPC)**

\[
\theta \left( \gamma - \frac{1}{2n} \right) < u \left( t^\Omega_R - t \left( \frac{1}{2} - \frac{1}{2n}, \gamma + \frac{1}{2n} \right) \right) - u(\pi + \eta) \\
= u \left( t \left( \gamma + \frac{1}{2n}, \frac{1}{2} - \frac{1}{2n} \right) - t^\Omega_L \right) - u(\pi + \eta),
\]

where \( t(s_L, s_R) \) is the Nash bargaining solution defined in Proposition 1. (iii). Note that the SPC depends on \( n \), both via the Nash bargaining solution and directly on the left-hand side. However, we have:

\[
\theta \left( \gamma - \frac{1}{2n} \right) < \theta \gamma \quad \text{and} \\
u \left( t^\Omega_R - t \left( \frac{1}{2}, \gamma \right) \right) < u \left( t^\Omega_R - t \left( \frac{1}{2} - \frac{1}{2n}, \gamma + \frac{1}{2n} \right) \right),
\]

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:

\[
\theta \gamma < u \left( t^\Omega_R - t \left( \frac{1}{2}, \gamma \right) \right) - u(\pi + \eta). \tag{13}
\]

We keep the \( n \)-dependent SPC as the main condition in the paper however, as the proof of the coming Proposition 2 in the Appendix shows that the \( n \)-dependent condition is both sufficient and necessary for \( \sigma^L \) and \( \sigma^R \) to not be equilibria of our stage 2 voting subgame if \( C = (E, E) \). Were we to use the (only sufficient) inequality in (13), it would be possible that even though (13) does *not* hold, \( \sigma^L \) and \( \sigma^R \) are still not equilibria of the subgame.\(^{19}\)

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.

**Proposition 2 (Nash equilibria of \( G - E \))**

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

\(^{19}\)That is, while \( \theta \left( \gamma - \frac{1}{2n} \right) \geq u \left( t^\Omega_R - t \left( \frac{1}{2} - \frac{1}{2n}, \gamma + \frac{1}{2n} \right) \right) - u(\pi + \eta) \Rightarrow \sigma^L, \sigma^R \) are equilibria, \( \theta \gamma \geq u \left( t^\Omega_R - t \left( \frac{1}{2}, \gamma \right) \right) - u(\pi + \eta) \) carries no such implication. When \( n \) is large enough however, it is clear that the strict inequality \( \theta \gamma > u \left( t^\Omega_R - t \left( \frac{1}{2}, \gamma \right) \right) - u(\pi + \eta) \) would be sufficient for \( \sigma^L \) and \( \sigma^R \) to be equilibria if \( C = (E, E) \).
(a) If \( C = (\emptyset, \emptyset) \), then \( \mathcal{E}^C = \{\sigma^L, \sigma^{LR}, \sigma^R\} \).

(b) If \( C = (E, \emptyset) \), then \( \mathcal{E}^C = \{\sigma^L, \sigma^R\} \).

(c) If \( C = (\emptyset, E) \), then \( \mathcal{E}^C = \{\sigma^L, \sigma^R\} \).

(d) If \( C = (E, E) \), then \( \mathcal{E}^C = \begin{cases} \{\sigma^{LR}\} & \text{if spc}, \\ \{\sigma^L, \sigma^{LR}, \sigma^R\} & \text{otherwise}. \end{cases} \)

The proof of Proposition 2 is provided in the Appendix. Proposition 2 reveals that—except when both conventional parties have credibly excluded the extreme party and (spc) holds—we encounter multiple equilibria. The election results immediately lead to possible government outcomes which we address in the next section.

6 Government Formation and Welfare

6.1 Possible Government Types

Here, we summarize the possible outcomes of government formation in the simplified “Core Game with PCs”, given some choices in the first stage. That is, we outline the equilibrium behavior in Stage 2, given any possible history in Stage 1. We denote the government formation by the parties in the government coalition. Propositions 1 and 2 immediately yield the possible forms of government, which we summarize in the next corollary:

Inspection of the proof also reveals that, if instead of assuming that \( E \) always enters into coalition with the conventional party with the most seats, we assume that it always enters into coalition with the party with the least seats, \( \sigma^{LR} \) will no longer be an equilibrium. Instead, a left-wing voter \( i \in \Omega^L \) would benefit by deviating to \( R \), so that \( LE \) occurs for sure, rather than with probability \( \frac{1}{2} \). One sees that instead \( \tilde{\sigma}^{LR} \) will be an equilibrium, in which every \( i \in \Omega^L \) votes \( R \) and every \( i \in \Omega^R \) votes \( L \). This is of course absurd, when confronted with reality. People never vote on purpose for a party they do not like because they believe it makes it more likely the party they like will be able to form a government. Moreover, in that case, it would no longer make sense to assume that voters which are not in the reduced game (i.e. \( i \notin \Omega \)) always vote for their preferred party, and this equilibrium also explicitly pits government formation objectives of voters against their disutility from voting dishonestly, making an explicit modelling of these conflicting objectives necessary. \( \sigma^L \) and \( \sigma^R \) would continue to be equilibria when \( C = (\emptyset, \emptyset) \) however, as long as the Strong PC held, although the inequalities would change slightly: the \( -\frac{\theta}{n} \) term would be replaced by \( -sE\theta \) and \( u(|t_i - t^L_E|) \) by \( u(|t_i - t^R_E|) \) in equation (19) for example. \( \sigma^R \) and \( \sigma^L \) would also continue to be the only equilibria when \( C = (E, \emptyset) \), but the relevant expected utility equations would also change. Our results regarding equilibrium selection in Sections 8.1 and 8.2 would change as well. Given the justifications we gave above for our assumption that the larger conventional party is able to enter government (and there are certainly more conceivable justifications), as well as the fact that the equilibrium \( \tilde{\sigma}^{LR} \) does not seem consistent with empirical observations, we believe our standing assumption is the more reasonable.
Corollary 2

(a) If $C = (\emptyset, \emptyset)$, then $(\sigma^L, L)$, $(\sigma^{LR}, LE)$, $(\sigma^{LR}, RE)$, or $(\sigma^R, R)$.

(b) If $C = (E, \emptyset)$, then $(\sigma^L, L)$ or $(\sigma^R, R)$.

(c) If $C = (\emptyset, E)$, then $(\sigma^L, L)$ or $(\sigma^R, R)$.

(d) If $C = (E, E)$, then $(\sigma^{LR}, LR)$ under (spc) and $(\sigma^L, L)$, $(\sigma^{LR}, LR)$ or $(\sigma^R, R)$ otherwise.

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

6.2 Welfare Ranking

In this subsection, we provide conditions on the 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 just the “swing voters” $\Omega$ of the “Core Game”. Through straightforward calculations, we find that

\[ W(t, d, s_G) = \bar{u}(t) - (1 - 2s_E) d - \theta s_G = \bar{u}(t) - 2\gamma d - \theta s_G. \quad (14) \]

Consider the following two conditions.

**Welfare Sensitivity-to-Perks Condition** (wspc): $\theta \gamma < \bar{u}(t_m) - \bar{u}(t_L)$.

Similarly to the spc introduced in Section 5, this requires that voters value perks less importantly than policy differences. Note that the two conditions are strictly different however, in the sense that one can find parameter configurations of the model where either holds, but not the other. Moreover wspc involves preferences of all voters $\mathcal{N}$ in the non-simplified game whereas spc only considers the two most extreme “swing
voters”. The following will make the SPC weaker, i.e. more general compared to the WSPC:

- more extreme swing voters, i.e. a large \( t^0_R \),
- less extreme parties i.e. a small \( t_R \) and large \( t_L \),
- a lower extreme party vote share, i.e. a small \( s_E \).

**Welfare Condition (wc):** \( 2\bar{d} \geq \theta \left( 1 + \frac{1}{2\gamma} \right) \).

This is the crucial assumption concerning the welfare effects of CPPs. Intuitively, it requires that conventional voters care more about avoiding the extreme party than about perks. The larger the share of the extreme party, that is, the smaller is \( \gamma \), the larger this preponderance of the disutility of the extreme policy relative to perks needs to be.

These conditions allow us to characterize the welfare ranking of the different equilibria as follows. If \( n \) is large enough means that there exists \( N \in \mathbb{N} \) such that the statement holds \( \forall n > N \). Also, we denote the policies resulting from a grand coalition where the left respectively the right party has one more seat than the other by

\[
\begin{align*}
\sigma^{LR}_{L} &:= t \left( \frac{1 - s_E}{2} + \frac{1}{2n} \right) - \frac{1}{2n} - \frac{1}{2n} , \\
\sigma^{LR}_{R} &:= 1 - \frac{1}{2n} + \frac{1 - s_E}{2} + \frac{1}{2n}.
\end{align*}
\]

**Proposition 3 (Welfare)**

(i) Given the “Core Game with PCs”, suppose in addition that all players \( i \in \mathcal{N} \) are symmetrically distributed around the median voter. If \( n \) is large enough, WSPC implies that

\[
(\sigma^{LR}, LR, t^L_{LR}) \sim_W (\sigma^{LR}, LR, t^R_{LR}) \succ_W (\sigma^L, L) \sim_W (\sigma^R, R),
\]

while \( \neg \)WSPC holding strictly implies the opposite strict welfare ordering.

(ii) If \( n \) is large enough, WC implies that

\[
(\sigma^{LR}, LR, t^L_{LR}) \sim_W (\sigma^{LR}, LR, t^R_{LR}) \succ_W (\sigma^{LR}, LE) \sim_W (\sigma^{LR}, RE),
\]

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

The proof of Proposition 3 can be found in the appendix.

---

\(^{21}\) That is, they will increase the right-hand side of the SPC compared to the right-hand side of the WSPC.
7 Equilibrium Selection

In this section, we lay out the foundations for the selection of possible equilibria. This is a central focus of the paper. For this purpose, it will be beneficial to introduce further notation. First, for each \( i \in \Omega \) and \( \sigma = (p_iLR)_{i \in \Omega} \), let \( \mathcal{X}^\sigma_i \) be the random variable that counts the number of votes for party \( L \) in \( \Omega \setminus \{i\} \), given that players different than \( i \) play according to \( \sigma_{-i} \). Second, let \( \sigma', \sigma'' \) and \( \sigma \) 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. We define the following function,

\[
\phi^C_i(\sigma'_i, \sigma''_i, \sigma_{-i}) := \mathbb{E}[U_i((\sigma'_i, \sigma_{-i})) | C] - \mathbb{E}[U_i((\sigma''_i, \sigma_{-i})) | C],
\]

which captures the preferences of voter \( i \) either for \( \sigma'_i \) or \( \sigma''_i \) depending on their beliefs about what the other voters decide. If \( \phi^C_i(\sigma'_i, \sigma''_i, \sigma_{-i}) \geq 0 \), then, given all other players decide on \( \sigma \), it is weakly better for player \( i \) to play \( \sigma'_i \) than \( \sigma''_i \). Third, recall that

\[
t(s_L, s_R) = \arg\max_{t \in [t_L, t_R]} s_L u(t - t_L) + s_R u(t_R - t)
\]

is the policy which is implemented when both conventional parties negotiate about their grand coalition platform.

Lemma 1
The following assertions hold for every \( i \in \Omega \):\(^{22}\)

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

\[
\phi^C_i(L, R, \sigma_{-i}) = \frac{1}{n} \theta \left[ \sum_{j=0}^{e-1} p(\mathcal{X}^\sigma_i = j) - \sum_{j=e+1}^{2e} p(\mathcal{X}^\sigma_i = j) \right] + p(\mathcal{X}^\sigma_i = e) \left[ u(|t^s_{LE} - t_i|) - u(|t^s_{RE} - t_i|) \right] + p(\mathcal{X}^\sigma_i = 0) \left[ u(|t^s_{RE} - t_i|) - u(|t_R - t_i|) - d \right] + p(\mathcal{X}^\sigma_i = 2e) \left[ u(|t_L - t_i|) - u(|t^s_{LE} - t_i|) + d \right].
\]

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

\[
\phi^C_i(L, R, \sigma_{-i}) = p(\mathcal{X}^\sigma_{-i} = 0) \left[ u(|t^w_{RE} - t_i|) - u(|t_R - t_i|) - d \right] + p(\mathcal{X}^\sigma_{-i} = 2e) \left[ u(|t_L - t_i|) - u(|t^w_{RE} - t_i|) + d - \left( s_E + \frac{1}{n} \right) \theta \right] + \frac{1}{n} \theta.
\]

\(^{22}\)Note 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 \( \phi^C_i \) to the proof itself.
(c) If $C = (\emptyset, E)$, then

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

$$
+ p(\mathcal{X}_{-i}^\sigma = 2e) \left[ u(|t_L - t_i|) - u(|t_{LE}^w - t_i|) + \bar{d} \right] - \frac{1}{n} \theta.
$$

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

$$
\phi_i^C(L, R, \sigma_{-i}) = p(\mathcal{X}_{-i}^\sigma = 0) \left[ u \left( \left| t \left( \gamma + \frac{1}{2n}, \frac{1}{2} - \frac{1}{2n} \right) - t_i \right| \right) 
$$

$$
- u(|t_R - t_i|) - \theta \left( \gamma - \frac{1}{2n} \right) \right] 
$$

$$
+ \sum_{j=1}^{2e-1} p(\mathcal{X}_{-i}^\sigma = j) \left[ u \left( \left| t \left( \gamma + \frac{1}{2n} + \frac{j}{n}, \frac{1}{2} - \frac{1}{2n} - \frac{j}{n} \right) - t_i \right| \right) 
$$

$$
- u \left( \left| t \left( \gamma - \frac{1}{2n} + \frac{j}{n}, \frac{1}{2} + \frac{1}{2n} - \frac{j}{n} \right) - t_i \right| \right) \right] 
$$

$$
+ p(\mathcal{X}_{-i}^\sigma = 2e) \left[ u(|t_L - t_i|) - u \left( \left| t \left( \frac{1}{2} - \frac{1}{2n}, \gamma + \frac{1}{2n} \right) - t_i \right| \right) 
$$

$$
+ \theta \left( \gamma - \frac{1}{2n} \right) \right].
$$

Before we proceed, we introduce further notation that will facilitate the presentation of the results. For every pair of profiles in pure strategies $\sigma_{-i}, \sigma'_{-i} \in \{L, R\}^{\Omega \setminus \{i\}}$ and for each $p \in [0, 1]$, we let $p\sigma_{-i} + (1 - p)\sigma'_{-i}$ denote a profile in mixed strategies where voter $j \in \Omega \setminus \{i\}$ votes according to $\sigma_{-i}$ with probability $p$ and according to $\sigma'_{-i}$ with probability $1 - p$.

### 7.1 Selection criteria

As already mentioned, multiplicity of equilibria constitutes a common problem in game theory. There are at least two ways to approach it. First, more strict conditions than that of Nash equilibrium can be imposed. If only one equilibrium survives the refinement, the multiplicity problem is solved. One such example is trembling-hand perfection (Selten, 1975), where each agent is assumed to “tremble” and play all actions with possible probability. Second, a (binary) dominance relation can be defined among different equilibria, e.g. risk-dominance (Harsanyi and Selten, 1988). On 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.\textsuperscript{23}

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 \textit{Domination in the presence of Uncorrelated Errors}, \textit{Domination in the presence of Correlated Errors}, \textit{Maximum Entropy in Strategies}, and \textit{Maximum Entropy in Outcomes}. We relate all definitions to the existing literature.

We note that 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 respectively with the strategy profiles where all players play the same strategy. Kim introduces four different criteria to support one equilibria over the other as the prediction 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 \(\phi_{i}^{C}\), defined in (15)\textsuperscript{24}:

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

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.

\textsuperscript{23}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.

\textsuperscript{24}Recall that \(C = (C_{L}, C_{R}) \in \{\emptyset, E\}^{2}\) is an arbitrary pure CPP strategy profile.
7.1.1 Coordination errors

Coordination trembling refers to the possibility that strategic voters may fail to coordinate on one equilibrium whenever there are multiple equilibria, even after having reached an agreement (e.g. through communication). There are at least two possibilities to incorporate potential failure in coordination on one equilibrium with respect to another equilibrium in the voting game. In both cases we compare equilibria of the voting game – given some CPP – by means of different binary relations.

On the one hand, we assume that there is a (common knowledge) probability that once agents have reached an agreement to coordinate on one equilibrium, the coordination fails resulting in all voters switching to coordinate on the other equilibrium. On the other hand, 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. Notice 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 $\varepsilon$-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”.

Let $\sigma' = (\sigma'_i)_{i \in \Omega}$, $\sigma'' = (\sigma''_i)_{i \in \Omega}$, and $\sigma = (\sigma_i)_{i \in \Omega}$ be three equilibria (which may include mixed strategies) of the (core) voting game for some given CPPs, say $C$. Note that by definition of the Nash equilibrium we have for every $i \in \Omega$:

$$\phi^C_i \left( \sigma'_i, \sigma''_i, \sigma_{-i} \right) \geq 0,$$

and that if $\sigma'_i = \sigma''_i$ then the definition of $\phi^C_i$ implies that

$$\phi^C_i \left( \sigma'_i, \sigma''_i, \sigma_{-i} \right) = 0.$$
Note also that by definition of $\phi^C_i$ in (15),

$$
\phi^C_i (\sigma'_i, \sigma''_i, \sigma_{-i}) = -\phi^C_i (\sigma''_i, \sigma'_i, \sigma_{-i}).
$$

(18)

Next, we formally introduce both definitions.

**Definition 1**

Given $C = (C_L, C_R) \in \{0, E\}^2$, we say that an equilibrium $\sigma'$ of the (core) voting subgame is immune to correlated errors at level $\varepsilon^*$ w.r.t. another equilibrium $\sigma''$ for player $i \in \Omega$ if

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

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

For a player $i$ that votes differently in $\sigma'$ and $\sigma''$, the above definition requires the expected value of playing their part in equilibrium $\sigma'$ not be lower than the expected value of playing their part in equilibrium $\sigma''$, given that there is a probability $1 - \varepsilon$ that all players play the agreed equilibria $\sigma'$ and a probability $\varepsilon$ that all players decide to vote according to $\sigma''$. Note that if $\sigma'_i = \sigma''_i$, by equation (17), we have

$$(1 - \varepsilon)\phi^C_i (\sigma'_i, \sigma''_i, \sigma_{-i}) + \varepsilon \phi^C_i (\sigma'_i, \sigma''_i, \sigma''_{-i}) = 0$$

automatically $\forall \varepsilon \in [0, 1]$. Also, notice that for fixed $\sigma', \sigma''$, $(1 - \varepsilon)\phi^C_i (\sigma'_i, \sigma''_i, \sigma_{-i}) + \varepsilon \phi^C_i (\sigma'_i, \sigma''_i, \sigma''_{-i})$ is a decreasing linear function of $\varepsilon$, thus in particular continuous in $\varepsilon$. This implies that in the definition of immunity to correlated errors above, it is equivalent to require $(1 - \varepsilon)\phi^C_i (\sigma'_i, \sigma''_i, \sigma_{-i}) + \varepsilon \phi^C_i (\sigma'_i, \sigma''_i, \sigma''_{-i}) \geq 0 \forall \varepsilon \in [0, \varepsilon^*]$ with $\varepsilon^*$ included. Indeed, if $\sigma'$ were immune to uncorrelated errors at level $\varepsilon^*$ w.r.t. $\sigma''$, but $(1 - \varepsilon^*)\phi^C_i (\sigma'_i, \sigma''_i, \sigma_{-i}) + \varepsilon^* \phi^C_i (\sigma'_i, \sigma''_i, \sigma''_{-i}) < 0$, by continuity, there would exist a small neighbourhood of $\varepsilon^*$ 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, since this is usually simpler.

We say that $\sigma'$ is immune to correlated errors at level $\varepsilon^*$ w.r.t. $\sigma''$ if it is so for each $i \in \Omega$.

\[\text{Note that when } \varepsilon^* = 0, \text{ the condition is always vacuously true, since any statement about the elements of the empty set is true. Also note that if } \sigma' \text{ is immune to correlated errors at level } \varepsilon^* \text{ w.r.t. } \sigma'', \text{ it is also so at any level } \varepsilon \leq \varepsilon^*.\]
Definition 2

Given \( C = (C_L, C_R) \in \{0, E\}^2 \), we say that an equilibrium \( \sigma' \) of the (core) voting subgame is immune to uncorrelated errors at level \( \varepsilon^* \) w.r.t. another equilibrium \( \sigma'' \) for player \( i \in \Omega \) if

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

where \( 0 \leq \varepsilon^* \leq 1. \)\(^{27}\)

For a player \( i \) that votes differently in \( \sigma' \) and \( \sigma'' \), the above definition requires that the expected value of playing their part in equilibrium \( \sigma' \) not be lower than the expected value of playing their part in equilibrium \( \sigma'' \), given that there is a probability \( 1 - \varepsilon \) that each player plays according to the agreed equilibrium \( \sigma' \) and a probability \( \varepsilon \) that they do not and vote according to \( \sigma'' \). Immunity at a level \( \varepsilon^* \) then means that if player \( i \) assumes that the errors only occur with probability at most \( \varepsilon^* \), it is still worth playing \( \sigma' \). If the voting game involves only two players, Definitions 1 and 2 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 \( \varepsilon \) for fixed \( \sigma', \sigma'' \). We thus provide the following Lemma, where we generalise our vocabulary to any abstract game.

Lemma 2

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 \( \sigma_i', \sigma''_i : X \to S_i \) and \( \sigma_{-i}', \sigma''_{-i} : X \to S_{-i} \) 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 \( (I\sigma_{-i}', I^c\sigma''_{-i}) : X \to S_{-i} \) denote the mixed strategy resulting from all players in \( I \) playing \( \sigma' \) and all players in \( I^c = \mathcal{N} \setminus (I \cup \{i\}) \) playing \( \sigma'' \). Suppose that randomization about which strategy \( \sigma' \) or \( \sigma'' \) 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 \( \sigma', \sigma'' \), and \( i \in \mathcal{N} \),

\(^{27}\)Note 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 \( \sigma' \) is immune to uncorrelated errors at level \( \varepsilon^* \) w.r.t. \( \sigma'' \), it is also so at any level \( \varepsilon \leq \varepsilon^* \).
the function
\[ \phi_i(\sigma'_i, \sigma''_i, (1 - \varepsilon)\sigma'_{-i} + \varepsilon\sigma''_{-i}) \]
\[ := \mathbb{E} \left[ U_i(\sigma'_i, (1 - \varepsilon)\sigma'_{-i} + \varepsilon\sigma''_{-i}) - U_i(\sigma''_i, (1 - \varepsilon)\sigma'_{-i} + \varepsilon\sigma''_{-i}) \right] \]
is a polynomial of order \(|N| - 1\) in \(\varepsilon\). In particular, it is continuous, has at most \(|N| - 1\) roots, and

\[ \exists B \subset \mathbb{R} \text{ with non-zero Lebesgue measure such that} \]
\[ \phi_i(\sigma'_i, \sigma''_i, (1 - \varepsilon)\sigma'_{-i} + \varepsilon\sigma''_{-i}) = 0 \quad \forall \varepsilon \in B \]
\[ \iff \sigma'_i = \sigma''_i. \]

Note that, as for the correlated case above, the fact that \(\phi_i^C(\sigma'_i, \sigma''_i, (1 - \varepsilon)\sigma'_{-i} + \varepsilon\sigma''_{-i})\) is continuous in \(\varepsilon\) allows us to conclude that \(\phi_i^C(\sigma'_i, \sigma''_i, (1 - \varepsilon)\sigma'_{-i} + \varepsilon\sigma''_{-i}) \geq 0 \forall \varepsilon \in [0, \varepsilon^*] \) is equivalent to \(\phi_i^C(\sigma'_i, \sigma''_i, (1 - \varepsilon)\sigma'_{-i} + \varepsilon\sigma''_{-i}) \geq 0 \forall \varepsilon \in [0, \varepsilon^*] \). In what follows we usually use the closed interval \([0, \varepsilon^*] \).

We say that \(\sigma'\) is immune to uncorrelated errors at level \(\varepsilon^*\) w.r.t. \(\sigma''\) if it is so for each \(i \in \Omega\).

In both Definition 1 and 2, we have introduced a binary relation between equilibria that need not be antisymmetric. In the following, we define antisymmetric relations using both definitions. We start with the “correlated errors” approach.

**Definition 3**

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

\[ \sigma'' \text{ is immune to correlated errors at level } \varepsilon^* \text{ w.r.t. } \sigma' \text{ for } i \implies \]
\[ \sigma' \text{ is immune to correlated errors at level } \varepsilon^* \text{ w.r.t. } \sigma'' \text{ for } i. \]

We use the notation \(\sigma' \succsim_i \sigma''\) to refer to this.

In words, Definition 3 requires that given some level of “correlated errors” \(\varepsilon^*\), if coordination in \(\sigma'\) fails w.r.t coordination in \(\sigma''\), then necessarily coordination in \(\sigma''\) fails w.r.t coordination in \(\sigma'\) at the same level \(\varepsilon^*\). We say that an equilibrium \(\sigma'\) strictly dominates an equilibrium \(\sigma''\) in the presence of correlated errors for player \(i \in \Omega\) if \(\sigma'\)
weakly dominates $\sigma''$ in the presence of correlated errors for $i$ and $\sigma'$ does not weakly dominate $\sigma''$ in the presence of correlated errors for $i$. We denote this by $\sigma' \succ \parallel_i \sigma''$.

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

The following lemma holds.

**Lemma 3**

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

(i) $\sigma'$ dominates $\sigma''$ in the presence of correlated errors for player $i$.

(ii) $\sigma'$ is immune to correlated errors at level $\varepsilon^* = \frac{1}{2}$ w.r.t. $\sigma''$ 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 in Appendix A also shows that (inequalities (iii) and (iv) hold strictly) $\Leftrightarrow (\sigma' \succ \parallel_i \sigma'') \Leftrightarrow (\exists \delta > 0$ such that $\sigma'$ is immune to correlated errors at level $\varepsilon^* = \frac{1}{2} + \delta$ w.r.t. $\sigma''$).

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

Next, we focus on the “uncorrelated errors” approach.
Definition 4
Given $C = (C_L, C_R) \in \{\emptyset, E\}^2$, we say that an equilibrium $\sigma'$ of the (core) voting subgame \textit{(weakly) dominates} an equilibrium $\sigma''$ in the presence of uncorrelated errors for player $i \in \Omega$ if, for each $\varepsilon^* \in (0, 1)$,

$\sigma''$ is immune to uncorrelated errors at level $\varepsilon^*$ w.r.t. $\sigma'$ for $i \implies$

$\sigma'$ is immune to uncorrelated errors at level $\varepsilon^*$ w.r.t. $\sigma''$ for $i$.

We use the notation $\sigma' \succ^{\perp}_i \sigma''$ to refer to this.

As before, we say that an equilibrium $\sigma'$ \textit{strictly dominates} an equilibrium $\sigma''$ in the presence of uncorrelated errors for player $i \in \Omega$ if $\sigma'$ weakly dominates $\sigma''$ in the presence of uncorrelated errors for $i$ and $\sigma''$ does not weakly dominate $\sigma'$ in the presence of uncorrelated errors for $i$. The extension to domination independently of player identity is analogous to the case of correlated errors. We use the notation $\succ^{\perp}$ 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 Lemma 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 $\varepsilon$. This is intuitive, since by Lemma 2 the function is a polynomial of order $n - 1$. One of the directions of Lemma 3 always holds. The following Lemma 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.

Lemma 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:

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

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

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

\[ \sigma' \text{ dominates } \sigma'' \text{ in the presence of uncorrelated errors for player } i \]

\[ \implies \sigma' \text{ is immune to uncorrelated errors at level } \varepsilon^* = \frac{1}{2} \text{ w.r.t. } \sigma'' \text{ for player } i. \]

Moreover, if \( \sigma' \succsim_{\varepsilon_i} \sigma'' \) then \( \exists \delta > 0 \) such that \( \sigma' \) is immune to uncorrelated errors at level \( \frac{1}{2} + \delta \) w.r.t. \( \sigma'' \) for \( i \).

The next Lemma provides a different sufficient condition for domination in the presence of uncorrelated errors.

**Lemma 5**

As before, let \( f_i(\varepsilon) := \phi_i^C(\sigma', \sigma'', (1 - \varepsilon)\sigma'_{-i} + \varepsilon\sigma''_{-i}) \). If \( f_i(\varepsilon) + f_i(1 - \varepsilon) \geq 0 \forall \varepsilon \in [0, 1] \), then \( \sigma' \) dominates \( \sigma'' \) in the presence of uncorrelated errors for player \( i \). If \( f_i(\varepsilon) + f_i(1 - \varepsilon) > 0 \), domination is strict.

Our final helping Lemma about equilibrium selection according to our uncorrelated errors criterion states that without further conditions, immunity at level \( \varepsilon^* = \frac{1}{2} \) is in fact not necessary for \( \sigma' \succsim \sigma'' \). In that sense it completes Lemma 4. We provide an example from our own (core) voting game.

**Lemma 6**

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

\[
\phi_i^{(\emptyset, \emptyset)}(\sigma^L_i, \sigma^R_i, (1 - \varepsilon)\sigma^L_{-i} + \varepsilon\sigma^R_{-i}) = \frac{\theta}{n} \left[ 2 \sum_{j=0}^{\varepsilon-1} p(B^l_{1-\varepsilon} = j) - 1 \right] \\
+ \binom{2e}{e} (1 - \varepsilon)^e \varepsilon^e \left[ u(t_i - t^*_L) - u(|t^*_R - t_i|) + \frac{\theta}{n} \right] \\
+ \varepsilon^{2e} \left[ u(|t^*_R - t_i|) - u(|t_R - t_i|) - \bar{d} \right] \\
+ (1 - \varepsilon)^{2e} \left[ u(t_i - t_L) - u(t_i - t^*_L) + \bar{d} \right].
\]

Define the three following quantities, which depend on the model parameters \( t_i, u(\cdot) \), \( t_L, t_R, t^*_L \) and \( t^*_R \), but not on \( \varepsilon \):

\[ K_1 := u(t_i - t^*_L) - u(|t^*_R - t_i|), \]
\[ K_2 := u(|t^*_R - t_i|) - u(|t_R - t_i|) - \bar{d}, \]
\[ K_3 := u(t_i - t_L) - u(t_i - t^*_L) + \bar{d}. \]
Note that by our assumptions about the policies, \( u(\cdot) \) and the strong \( \text{pc} \) (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, \( \sigma^R >_i ^{-} \sigma^L \), for all \( i \in \Omega^R \) and all values of \( K_1, K_2, K_3, n, e \) and \( \theta \). In addition, we provide four examples of parameter vectors \( (n, e, \theta, K_1, K_2, K_3) \) where \( \sigma^R \) is not immune to uncorrelated errors at level \( \frac{1}{2} \) w.r.t. \( \sigma^L \).

Lemma 6 proves that no exact equivalent of Lemma 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 \( \sigma^R \) is not immune to uncorrelated errors at level \( \frac{1}{2} \) w.r.t. \( \sigma^L \). 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.

7.1.2 Entropy in equilibria and strategies

Our approach in this section builds on the concept of risk-dominance (Harsanyi and Selten, 1988). Let us introduce the latter notion for coordination games with player set \( N = \{1, 2\} \), 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 \( u_2(B, B) > u_2(B, A) \). Then, \( (A, A) \) 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 \( u_2(\frac{1}{2}A + \frac{1}{2}B, A) > u_1(\frac{1}{2}A + \frac{1}{2}B, B) \).

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

Next, we present several extensions of the above notion.
Definition 5
We say that an equilibrium of the (core) voting subgame $\sigma'$ (weakly) dominates another equilibrium $\sigma''$ according to maximum entropy in equilibria for player $i \in \Omega$, if
\[
\frac{1}{2} \phi_i^C (\sigma_i', \sigma_i'', \sigma_{-i}') + \frac{1}{2} \phi_i^C (\sigma_i', \sigma_i'', \sigma_{-i}'') \geq 0.
\]

Strict domination for player $i \in \Omega$ is defined naturally with a strict inequality. We say that $\sigma'$ (weakly) dominates $\sigma''$ according to maximum entropy in equilibria, if it does so for every player $i \in \Omega$. 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 $\sigma'$ dominates $\sigma''$ according to maximum entropy in equilibria for player $i \in \Omega$ if and only if $\sigma'$ dominates $\sigma''$ 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.

Definition 6
We say that an equilibrium of the (core) voting subgame $\sigma'$ weakly dominates another equilibrium $\sigma''$ according to maximum entropy in strategies (version 1) for player $i \in \Omega$, if
\[
\phi_i^C (\sigma_i', \sigma_i'', \frac{1}{2} \sigma_{-i}' + \frac{1}{2} \sigma_{-i}'') \geq 0.
\]

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 Lemma 4, it follows that if $\phi_i^C (\sigma_i', \sigma_i'', (1 - \varepsilon) \sigma_{-i}' + \varepsilon \sigma_{-i}'')$ 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 6 shows, the “near-monotonicity” property is by no-means assured. Thus, it is likely 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. Unfortunately, in our voting game, no such examples exist.
Definition 7
We say that an equilibrium of the (core) voting subgame $\sigma'$ dominates another equilibrium $\sigma''$ according to maximum entropy in strategies (version 2) for player $i \in \Omega$, if
\[
\phi_i^C \left( \sigma_i', \sigma_i'', \frac{1}{2} \sigma_{-i}^L + \frac{1}{2} \sigma_{-i}^R \right) \geq 0.
\]
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 $\frac{1}{2}$.

8 Equilibrium Selection for the Voting Subgame

8.1 Equilibria without Credible Exclusion Promises

After laying the foundations regarding equilibrium selection in the previous section, we now start by discussing the selection of equilibria without CPPs. Hence, let $C = (\emptyset, \emptyset)$. According to Proposition 2, we have $\mathcal{E}^C = \{\sigma^L, \sigma^{LR}, \sigma^R\}$. We make pairwise comparisons among all equilibria according to all criteria except 'entropy in strategies (version 2)'. When we say that a statement holds “if $n$ and $e$ large enough”, we mean that given all other model parameters, there exists an $N \in \mathbb{N}$ such that for all $n > N$ and $e \geq e(n) := \left\lfloor \frac{\tilde{s}_E}{2} \right\rfloor$, the statement holds, where $\tilde{s}_E$ is an “ideal parliamentary seat share” for the extreme party. That is, we make the following asymptotic assumption: $e$ increases at least as fast as $e(n)$, which stays at an approximately constant ratio to $n$, such that the extreme party seat share is always below but close to the “ideal parliamentary seat share”.

Proposition 4
(i) $\sigma^L, \sigma^R \succ \parallel \sigma^{LR}$, that is $\sigma^L$ and $\sigma^R$ strictly dominate $\sigma^{LR}$ in the presence of correlated errors, if $n \geq 9$ and $e \geq 2$.

(ii) If $n$ and $e$ are large enough, $\sigma^{LR}$ is immune to uncorrelated errors at level $\frac{1}{2}$ w.r.t. both $\sigma^L$ and $\sigma^R$. In particular, $\sigma^{LR} \succ \perp \sigma^L, \sigma^R$, that is $\sigma^{LR}$ strictly dominates $\sigma^L$ and $\sigma^R$ in the presence of uncorrelated errors. Moreover, given $\bar{t} \in \left(\frac{1}{2}, 1\right]$ (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 $\sigma^{LR}$ to $\sigma^L$, 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 $\sigma^R$.

(iii) $\sigma^L$ and $\sigma^R$ strictly dominate $\sigma^{LR}$ according to maximum entropy in equilibria, if $n \geq 9$ and $e \geq 2$.

(iv) If $n$ and $e$ are large enough, $\sigma^{LR}$ strictly dominates $\sigma^L$ and $\sigma^R$ according to maximum entropy in strategies (version 1).

(v) $\sigma^L \succ^i \sigma^R \forall i \in \Omega^L$ and $\sigma^R \succ^i \sigma^L \forall i \in \Omega^R$. Thus in particular, $\sigma^L$ and $\sigma^R$ are not comparable according to the ordering given by the correlated errors criterion (neither dominates the other).

(vi) $\sigma^L \succ^i \perp \sigma^R \forall i \in \Omega^L$ and $\sigma^R \succ^i \perp \sigma^L \forall i \in \Omega^R$. Thus in particular, $\sigma^L$ and $\sigma^R$ are not comparable according to the ordering given by the uncorrelated errors criterion (neither dominates the other).

(vii) $\sigma^L$ and $\sigma^R$ are not comparable according to the ordering given by the maximum entropy in equilibria criterion (neither dominates the other).

(viii) $\sigma^L$ and $\sigma^R$ are not comparable according to the ordering given by both versions (they are identical in this case) of the maximum entropy in strategies criterion (neither dominates the other).

8.2 Equilibria with Credible Exclusion of the Extreme party

In this subsection, we assume that during campaigns, the conventional parties can commit to not forming a coalition with the extreme party to govern the country. In Section 10 we will detail which type of commitment devices might be available. In this section, we assume that these devices are indeed available. We use the definitions of coordination/selection criteria laid out in sections 7.1.1 and 7.1.2 and, for some statements, introduce the following condition:

**Preclusion-Promises Condition (PPC):** $s_E \theta > u(|\eta|) - u(\pi + \eta)$.

Let $C = (E, \emptyset)$. According to Proposition 2, we have $\mathcal{E}^C = \{\sigma^L, \sigma^R\}$. We compare these two equilibria according to all criteria. The meaning of “if $n$ and $e$ large enough”
is identical to what it was in Proposition 4. In particular, recall that this implies that $e$ and $n$ grow at an approximately constant ratio.

**Proposition 5**

(i) If $n$ is large enough (irrespective of $e$), $\sigma^R \succ_i \sigma^L \ \forall i \in \Omega^R \cup \{m\}$. If in addition PPC holds, $\sigma^R \succ_i \sigma^L$ also $\forall i \in \Omega^L$. In that case $\sigma^R$ strictly dominates $\sigma^L$ in the presence of correlated errors.

(ii) If $n$ and $e$ are large enough, $\sigma^L$ is immune to uncorrelated errors at level $\frac{1}{2}$ w.r.t. $\sigma^R$, and in particular $\sigma^L \succ^\perp \sigma^R$, that is $\sigma^L$ strictly dominates $\sigma^R$ in the presence of uncorrelated errors. Moreover, if $n$ and $e$ are large enough$^{28}$, for those players with $t_i < t^w_{RE}$, the 'entropy in strategies' and 'uncorrelated errors' criteria are always equivalent, while if the Weak PC holds strictly, this equivalence holds for all players $i \in \Omega$.

(iii) If $n$ is large enough (irrespective of $e$), $\sigma^R$ strictly dominates $\sigma^L$ according to maximum entropy in equilibria $\forall i \in \Omega^R \cup \{m\}$. If in addition PPC holds, then $\sigma^R$ also strictly dominates $\sigma^L$ according to maximum entropy in equilibria $\forall i \in \Omega^L$.

(iv) If $n$ and $e$ are large enough, $\sigma^L$ strictly dominates $\sigma^R$ according to maximum entropy in strategies (version 1 and version 2, as these are equivalent here).

Note that by symmetry, the exact same statements, *mutatis mutandis* will hold when $C = (\emptyset, E)$.

Proposition 5 reveals that under certain conditions, exclusion of the extreme party by one party will cause the equilibrium where voters coordinate on this party to be more stable than the equilibrium where voters coordinate on the non-excluding conventional party. This may lead it to be selected in the long run.

9 Solving the Full Game when CPPs are Credible

We next investigate whether promises are made in equilibrium, assuming that CPPs are available, that is, we now assume that the parties are able to endogenously select which CPP $\in \{\emptyset, E\}$ they wish to make. In choosing whether to make CPPs, the conventional parties take into account which equilibria will result in the different subgames after having decided on their CPPs.

$^{28}$They may possibly need to be larger than for the first statement.
Suppose \( C = (\emptyset, \emptyset) \) results in \( \sigma^{LR} \) being played and hence in a government of \((LE)\) or \((RE)\) each with probability \( \frac{1}{2} \); \( C = (E, \emptyset) \) triggers \( \sigma^L \) and thus \((L)\); and \( C = (\emptyset, E) \) triggers \( \sigma^R \) and \((R)\). Additionally assuming that the \( spc \) holds, and hence if \( C = (E, E) \), the unique resulting equilibrium is \( \sigma^{LR} \) with \((LR)\), we obtain the following choice game in Stage 1 as displayed in Table 2.

<table>
<thead>
<tr>
<th>Party ( L )</th>
<th>( \emptyset )</th>
<th>( E )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \emptyset )</td>
<td>( \frac{1}{2} V_L (t_{LE}^s, \bar{d}, 1) + \frac{1}{2} V_L (t_{RE}^s, \bar{d}, 0) ), ( \frac{1}{2} V_R (t_{LE}^s, \bar{d}, 0) + \frac{1}{2} V_R (t_{RE}^s, \bar{d}, 1) )</td>
<td>( V_L (t_R, 0, 0) ), ( V_R (t_R, 0, 1) )</td>
</tr>
<tr>
<td>( E )</td>
<td>( V_L (t_L, 0, 1) ), ( V_R (t_L, 0, 0) )</td>
<td>( \frac{1}{2} V_L (t_{LR}^s, 0, \bar{d}, 0) + \frac{1}{2} V_L (t_{LR}^R, 0, \bar{d}, 1) ), ( \frac{1}{2} V_R (t_{LR}^s, 0, \bar{d}, 0) + \frac{1}{2} V_R (t_{LR}^R, 0, \bar{d}, 1) )</td>
</tr>
</tbody>
</table>

Table 2: Conventional parties’ contract choice game

Here, as in Section 6,

\[
t_{LR}^L := t \left( \frac{1 - s_E}{2} + \frac{1}{2n}, \frac{1 - s_E}{2} - \frac{1}{2n} \right)
\]

\[
t_{LR}^R := 1 - t_{LR}^L = t \left( \frac{1 - s_E}{2} - \frac{1}{2n}, \frac{1 - s_E}{2} + \frac{1}{2n} \right)
\]

It follows immediately that \( C = (E, E) \) is the only equilibrium of the game. In this situation, CPPs are most effective in preventing extreme parties from entering government and ensuring moderate policy. Our previous results in Propositions 4 and 5 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 6**

Given the “Core Game with \( pc \)”, suppose that \( spc \) holds, \( n \) 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 equilibrium of the political game is:

- Both parties make CPPs excluding the extreme party \( E \), \( C = \{E, E\} \).

---

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.
- Voters vote sincerely according to $\sigma^{LR}$.

- Both conventional parties will form a grand coalition implementing policy $(LR)$.

Proposition 6 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 5. (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 4. (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. However, we also discussed the case where errors, respectively deviations from the coordination equilibrium may be correlated. Then, Proposition 5 predicts that the unilateral CPP using party could not expect higher but lower voter support and hence CPPs would not be used. It is interesting, however, that according to the same criteria Proposition 4 would not predict the extreme party entering government either but rather a single party government by one of the conventional parties in the case when no CPPs are made. Consequently, when these selection criteria are applied CPPs are not necessary to prevent the extremist party from entering government.

While in the latter case, the selection criteria would predict a single-party government,
in the case with uncorrelated errors, respectively deviations, a grand coalition will be the predicted outcome.

In terms of welfare, we summarize our results with the following Corollary.

**Corollary 3 (Welfare Consequences of Credible CPPs)**

Given the “Core Game with pcs”, suppose that spc and wc hold, n and e are large enough, and equilibria are selected according to one of the four criteria we introduced in section 7. Then the possibility to use (credible) CPPs is weakly welfare improving. If equilibria are selected according to “domination in the presence of uncorrelated errors” or “maximum entropy in strategies”, CPPs are strictly welfare improving, while 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.

Proposition 3 shows that which of these two cases – coordination on either conventional party, or a grand coalition – yields higher welfare, depends on the level of perks relative to the polarization of political platforms, that is, on whether wspc holds or not.

## 10 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 his promises, as long as he has 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 he delivers 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.

11 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) 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 predict 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.
Appendix

A Proofs

Proof of Corollary 1 (Polarization)

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

Case: $i \in \Omega^R \cup \{\frac{1}{2}\}$, that is $t_i \in [\frac{1}{2}, t^j_R].$

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

$$\frac{\partial}{\partial t_i} \left\{ u(|t^j_{RE} - t_i|) - u(t_i - t_L) \right\} = -u'(t^j_{RE} - t_i) - u'(t_i - t_L) > 0$$

since $u'(x) < 0 \forall x > 0$ is assumed. On the other hand, if $t_i > t^j_{RE},$ we have

$$\frac{\partial}{\partial t_i} \left\{ u(|t^j_{RE} - t_i|) - u(t_i - t_L) \right\} = u'(t^j_{RE} - t_i) - u'(t_i - t_L) \geq 0$$

because $u'$ is decreasing ($u$ is assumed to be (weakly) concave).\(^{33}\) Now the fact that $u(|t^j_{RE} - t_i|) - u(t_i - t_L)$ is increasing in $t_i$ implies that

$$u(|t^j_{RE} - t^j_R|) - u(t^j_R - t_L) \geq u(|t^j_{RE} - t^j_R|) - u(t^j_R - t_L) \quad \forall t_i \in [\frac{1}{2}, t^j_R]$$

So that $(iv)$ is implied by

$$d \geq s_E \theta + u(|t^j_{RE} - t^j_R|) - u(t^j_R - t_L)$$

which are precisely the PCs.

\(^{33}\)Technically, $u(x)$ is not differentiable in the classic two-sided limit sense, at $x = 0$. Thus, we should treat the cases $t_i = t_L = t^j_{RE} = \frac{1}{2}$ (recall that inequality (4) is assumed) and $t_i = t^j_{RE}$ separately. However, the fact that $u(|t^j_{RE} - t_i|) - u(t_i - t_L)$ is increasing on $[\frac{1}{2}, t^j_R]$ follows from the positive derivative via the mean-value theorem, which for the case $t^j_{RE} \neq \frac{1}{2}$ would have to be applied twice, once on the subdomain $[\frac{1}{2}, t^j_{RE}]$ and once on the subdomain $[t^j_{RE}, t^j_R].$ 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.
Secondly, consider the set of inequalities with a negative “−” sign, that is (i) is the version with \( d \geq s_E \theta - \left( u(t^{j^i}_{LE} - t_i) - u(t_R - t_i) \right) \) 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 \(-u(t^{j^i}_{LE} - t_i) + u(t_R - t_i)\) is increasing in \( t_i \) on \( \left[ \frac{1}{2}, t^\alpha_R \right] \). Indeed, if \( t_i < t_R \), we have

\[
\frac{\partial}{\partial t_i} \left\{ u(t_R - t_i) - u(t_i - t^{j^i}_{LE}) \right\} = -u'(t_R - t_i) - u'(t_i - t^{j^i}_{LE}) > 0
\]

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

\[
\frac{\partial}{\partial t_i} \left\{ u(t_R - t_i) - u(t_i - t^{j^i}_{LE}) \right\} = u'(t_i - t_R) - u'(t_i - t^{j^i}_{LE}) \geq 0
\]

because \( u' \) is (weakly) decreasing. Now as before, the fact that \( u(t_R - t_i) - u(t_i - t^{j^i}_{LE}) \) is increasing in \( t_i \) implies that

\[
u(t_R - t^{\alpha}_{LE}) - u(t^{\alpha}_{LE} - t^{j^i}_{LE}) = u(\eta) - u(\pi + \eta - \beta^i) \geq u(t_R - t_i) - u(t_i - t^{j^i}_{LE}) \ \ \forall t_i \in \left[ \frac{1}{2}, t^\alpha_R \right]
\]

so that (i) is implied by

\[
d \geq s_E \theta + u(\eta) - u(\pi + \eta - \beta^i)
\]

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

\[
g(\alpha) = u(|\alpha \beta^i + \eta|) - u(\pi + \eta + (\alpha - 1)\beta^i).
\]

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 \( \alpha \). Indeed, first assume that \( \eta \geq 0 \). Then

\[
g'(\alpha) = \beta^i u'(\alpha \beta^i + \eta) - \beta^i u'(\pi + \eta + (\alpha - 1)\beta^i)
\]

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

\[
g'(\alpha) \geq 0 \iff \pi + \eta + (\alpha - 1)\beta^i \geq \alpha \beta^i + \eta \ \ \forall \alpha \in [0, 1].
\]

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

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

\[
g'(\alpha) = -\beta^i u'(-\alpha \beta^i - \eta) - \beta^i u'(\pi + \eta + (\alpha - 1)\beta^i) > 0
\]
since $u'(x) < 0 \ \forall x > 0$, and the claim is proved again by the mean-value theorem.

Case: $i \in \Omega^L$, that is $t_i \in \left[ t_i^\Omega, \frac{1}{2} \right)$.

Again, consider first 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 $\left[ t_i^\Omega, \frac{1}{2} \right)$. 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

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

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 $\left[ t_i^\Omega, \frac{1}{2} \right)$. This implies that it is sufficient to show

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

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 2 (Nash equilibria of $G^{-E}$)

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

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

First, we prove that $\sigma^{LR}$ is an equilibrium. Indeed, according to Proposition 1, if voters vote according to $\sigma^{LR}$, the implemented policy is either $(t^*_{LE}, \tilde{d})$ or $(t^*_{RE}, \tilde{d})$ 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 $(t^*_{RE}, \tilde{d})$ for sure with support $s_G^i = s_G \pm \frac{1}{n}$, while if $i \in \Omega^R$ deviates and votes strategically for $L$, the implemented policy is $(t^*_{LE}, \tilde{d})$ for sure with support $s_G^i = s_G \pm \frac{1}{n}$. Then, for each $i \in \Omega^L$,

$$E \left[ U_i \left( (L, \sigma^{LR}_{-i}) \right) \right] - E \left[ U_i \left( (R, \sigma^{LR}_{-i}) \right) \right]$$

$$= \frac{u(|t_i - t^*_{LE}|) + u(|t_i - t^*_{RE}|)}{2} - \tilde{d} - \theta s_G - \frac{\theta}{2n} > 0,$$

since by assumption $u$ is decreasing, $t_i < \frac{1}{2}$ implies $|t_i - t^*_{LE}| \geq t^*_{RE} - t_i$ and $\theta > 0$.

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

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

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

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

As a consequence, $\sigma^{LR}$ is a (weak\textsuperscript{34}) Nash equilibrium of the voting subgame.

Second, we prove that $\sigma^L$ is an equilibrium if the Strong PC holds. According to Proposition 1, if voters vote according to $\sigma^L$, 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

\textsuperscript{34}Because of the median voter.
(\(t^*_{LE}, \bar{d}\)) with support \(s_G = \frac{1}{2} - \frac{1}{2n}\). Then, for each \(i \in \Omega\),

\[
\mathbb{E}\left[U_i\left((L, \sigma^L_{-i})\right)\right] - \mathbb{E}\left[U_i\left((R, \sigma^L_{-i})\right)\right] \\
= u(|t_i - t_L|) - \theta \left(\frac{1}{2} + \frac{1}{2n}\right) - u(|t_i - t^*_{LE}|) - \bar{d} - \theta \left(\frac{1}{2} - \frac{1}{2n}\right) \\
= u(|t_i - t_L|) - u(|t_i - t^*_{LE}|) + \bar{d} - \frac{\theta}{n}.
\]

We see that if

\[
\bar{d} \geq \frac{\theta}{n} + u(|t_i - t^*_{LE}|) - u(|t_i - t_L|) \quad \forall i \in \Omega,
\]

then it is not profitable for any \(i \in \Omega\) to deviate and \(\sigma^L\) 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 Corollary 1.(ii).\(^{35}\) Hence \(\sigma^L\) is a strict Nash equilibrium.

Third, we focus on \(\sigma^R\). According to Proposition 1, if voters vote according to \(\sigma^R\), 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 \((t^*_{RE}, \bar{d})\) with support \(s_G = \frac{1}{2} - \frac{1}{2n}\). Analogously to the analysis for \(\sigma^L\), we can show that \(\sigma^R\) is an equilibrium if

\[
\bar{d} \geq \frac{\theta}{n} + u(|t_i - t^*_{RE}|) - u(|t_i - t_R|) \quad \forall i \in \Omega
\]

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

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

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

\[
\mathbb{E}\left[U_i\left((R, \sigma^{LR}_{-i})\right)\right] - \mathbb{E}\left[U_i\left((L, \sigma^{LR}_{-i})\right)\right] \\
= u(|t_i - t^w_{RE}|) - \bar{d} - \theta \left(\frac{1 - \frac{s_E}{2}}{2}\right) - u(|t_i - t^w_{RE}|) - \bar{d} - \theta \left(\frac{1 - \frac{s_E}{2}}{2} - \frac{1}{n}\right) \\
= -\frac{\theta}{n} < 0.
\]

\(^{35}\)It can be shown in a similar fashion to the proof of Corollary 1 that \(u(|t_i - t^*_{LE}|) - u(|t_i - t_L|)\) is increasing in \(t_i\) for all \(t_i \in [t^\Omega_{LE}, t^\Omega_R]\). Hence, one can show that \(\exists n \in \mathbb{N}\) such that \(\sigma^L\) is a Nash equilibrium if and only if \(\bar{d} > u(|t^\Omega_{R} - t^*_{LE}|) - u(|t^\Omega_R - t_L|) = u(\pi + \eta - \beta^*) - u(\pi + \eta)\).
Hence, $\sigma^{LR}$ is not a Nash equilibrium of the voting subgame.

Second, we prove that $\sigma^L$ is an equilibrium if the Weak pc holds. Indeed, according to Proposition 1, if voters vote according to $\sigma^L$, 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 $i \in \Omega$,

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

by Corollary 1.(iv). Note that we need this inequality to hold also for $t_i = t_R^0$, and thus, $\sigma^L$ is a strict Nash equilibrium if and only if the Weak pc holds strictly.

Third, we study when $\sigma^R$ is an equilibrium. According to Proposition 1, if voters vote according to $\sigma^R$, 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 $\sigma^L$ case, it can be checked that $\sigma^R$ 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 Corollary 1.(iii) and the fact that $e \geq 1$ is assumed.

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

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

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

which again holds strictly (and hence $\sigma^L$ is a strict equilibrium) as long as the the Weak pc holds; and that $\sigma^R$ is an equilibrium if

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

which is implied by Corollary 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 $\sigma^{LR}$ is an equilibrium. Indeed, let $s^{LR} = \frac{1-s_E-\frac{1}{2}}{2}$, and recall the definition of

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

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

$$t(s_L) := t(s_L, 1 - s_E - s_L).$$

According to Proposition 1, if voters vote according to $\sigma^{LR}$, the implemented policies are $(t(s^{LR} + \frac{1}{n}), 0)$ or $(t(s^{LR} - \frac{1}{n}), 0)$, each with probability $\frac{1}{2}$ and support $s_G = 1 - s_E$. If $i \in \Omega^L$ deviates and votes strategically for $R$, 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 $s_G = 1 - s_E$. 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 $t_L \neq t_R^{36}$, we have:

$$t(s_L) \text{ is strictly decreasing in } s_L \tag{22}$$

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

$$t\left(s^{LR} + \frac{1}{n}\right) + t\left(s^{LR}\right) = 1 \tag{23}$$

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

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

$$= \left[u\left(|t_i - t(s^{LR} + \frac{1}{n})|\right) + u\left(|t_i - t(s^{LR})|\right) - \theta s_G\right]$$

$$- \left[u\left(|t_i - t(s^{LR})|\right) + u\left(|t_i - t(s^{LR} - \frac{1}{n})|\right) - \theta s_G\right]$$

$$= \frac{u\left(|t_i - t(s^{LR} + \frac{1}{n})|\right) - u\left(|t_i - t(s^{LR} - \frac{1}{n})|\right)}{2} > 0,$$

$^{36}$If $t_L = t_R = \frac{1}{2}$, then of course $t(s_L, s_R) = t_m = \frac{1}{2}$ whatever the vote shares $s_L$ and $s_R$. 50
because \( t_i < \frac{1}{2} \) and \( t \left( s^{LR} + \frac{1}{n} \right) < t \left( s^{LR} - \frac{1}{n} \right) \), by (22).

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 (23)), the median voter, \( m \), is indifferent between voting for either party. As a consequence, \( \sigma^{LR} \) is a (weak) Nash equilibrium of the voting subgame.

Second, we consider \( \sigma^L \) and show that this is not an equilibrium if and only if the SPC holds. Indeed, according to Proposition 1, if voters vote according to \( \sigma^L \), 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 \( \left( t \left( s^L_G - \frac{1}{n} \right), 0 \right) \) with support \( s^i_G = 1 - s_E \). Note that \( t_L < t \left( s_G - \frac{1}{n} \right) \). Then, for each \( i \in \Omega \),

\[
\mathbb{E} \left[ U_i \left( (L, \sigma^{-i}_{LR}) \right) \right] - \mathbb{E} \left[ U_i \left( (R, \sigma^{-i}_{LR}) \right) \right] = u^i \left( t \right) - u^i \left( t \left( s^i_G - \frac{1}{n} \right) \right) = 0,
\]

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,

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

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

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

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

\[
\frac{\partial}{\partial t_i} \left\{ u \left( |t_i - t_L| \right) - u \left( |t_i - t \left( s_G - \frac{1}{n} \right)\right| \right) \right\} = u'(t_i - t_L) - u' \left( t_i - t \left( s_G - \frac{1}{n} \right) \right) t_L \leq 0,
\]
where again the inequality holds since $u(\cdot)$ is (weakly) concave. As a result, by the mean-value theorem, we have:

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

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

$$\theta \left(\gamma - \frac{1}{2n}\right) \geq u\left(|t_R^\Omega - t \left(s_G - \frac{1}{n}\right)\right) - u(|t_R^\Omega - t_L|), \quad (24)$$
which is precisely the inverse inequality appearing in the SPC.

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

$$\theta \left(\gamma - \frac{1}{2n}\right) \geq u\left(|t_L^\Omega - t \left(\gamma + \frac{1}{2n}\right)\right) - u(|t_L^\Omega - t_R|). \quad (25)$$

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 (25) is identical to the right-hand side of (24) and both are precisely the inverse of the SPC.
Proof of Proposition 3

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

$$s_{(\sigma^L,L)} = s_{(\sigma^R,R)} = \frac{1}{2} + \frac{1}{2n}$$

$$s_{(\sigma^{LR},LR)} = 1 - s_E$$

$$s_{(\sigma^{LR},LE)} = s_{(\sigma^{LR},RE)} = \frac{1}{2}(1 - s_E) + \frac{1}{2n}.$$ 

Proof (i).

By the formula in (14), we have:

$$W\left(t_L, 0, \frac{1}{2} + \frac{1}{2n}\right) = \bar{u}(t_L) - \theta \left(\frac{1}{2} + \frac{1}{2n}\right)$$

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

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

$$W\left(t^{LR}_{LR}, 0, 1 - s_E\right) = W\left(t^R_{LR}, 0, 1 - s_E\right)$$

$$= \bar{u}(t^L_{LR}) - \theta(1 - s_E),$$

and we see that

$$W\left(\Phi(\sigma^{LR}, LR, t^L_{LR})\right) = W\left(t^L_{LR}, 0, 1 - s_E\right) = W\left(\Phi(\sigma^{LR}, LR, t^R_{LR})\right)$$

$$> W\left(\Phi(\sigma^L, L)\right) = W\left(t_L, 0, \frac{1}{2} + \frac{1}{2n}\right) = W\left(\Phi(\sigma^R, R)\right)$$

$$\Leftrightarrow \theta \left(\gamma - \frac{1}{2n}\right) < \bar{u}(t^L_{LR}) - \bar{u}(t_L), \quad (26)$$

where $\Phi$ 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 (26), 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

$$\theta \gamma > \theta \left(\gamma - \frac{1}{2n}\right), \quad \text{and}$$

$$\bar{u}(t_m) - \bar{u}(t_L) > \bar{u}(t^L_{LR}) - \bar{u}(t_L),$$
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 $n$,

$$\lim_{n \to \infty} t_{LR}^L(n) = \lim_{n \to \infty} t \left( \frac{1-s_E}{2} + \frac{1}{2n}, \frac{1-s_E}{2} - \frac{1}{2n} \right) = t_m = \frac{1}{2},$$

and

$$\lim_{n \to \infty} \left\{ \gamma - \frac{1}{2n} \right\} = \gamma.$$ 

If $n$ is large enough, the left-hand and right-hand sides of (26) can both be made to be contained in disjoint $\varepsilon$-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:

$$W \left( t_{LE}^j, \bar{d}, \frac{1-s_E}{2} + \frac{1}{2n} \right) = W \left( t_{RE}^j, \bar{d}, \frac{1-s_E}{2} + \frac{1}{2n} \right)$$

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

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

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

(27)

However, as long as $t_{LE}^s < \frac{1}{2},^{37}$ in the limit

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

and so inequality (27) 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 $n$.

---

$^{37}$Recall that $t_{LE}^w \leq t_{LE}^s$, by assumption.
Proof of Lemma 1

Let $\sigma$ be an arbitrary strategy profile. We distinguish four cases depending on the CPPs written in Stage 1. We build on Proposition 1 and the discussion in subsection 5.1 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 $i$, $\chi_{\sigma}^{\sigma_i}$.

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

On the one hand,

$$
\mathbb{E} [U_i((L, \sigma_{-i}))] = p(X_{\sigma_i}^{\emptyset} = 0) \left[ u(|t_{RE}^s - t_i|) - \bar{d} - \theta \left( \frac{1}{2} - \frac{1}{2n} \right) \right] + \sum_{j=1}^{e-1} p(X_{\sigma_i}^{\emptyset} = j) \left[ u(|t_{RE}^s - t_i|) - \bar{d} - \theta \left( \frac{1}{2} - \frac{1}{2n} - \frac{j}{n} \right) \right] + p(X_{\sigma_i}^{\emptyset} = e) \left[ u(|t_{LE}^s - t_i|) - \bar{d} - \theta \left( \frac{1}{2} (1 - s_E) + \frac{1}{2n} + \frac{j - e}{n} \right) \right] + p(X_{\sigma_i}^{\emptyset} = 2e) \left[ u(|t_L - t_i|) - \theta \left( \frac{1}{2} + \frac{1}{2n} \right) \right].
$$

On the other hand,

$$
\mathbb{E} [U_i((R, \sigma_{-i}))] = p(X_{\sigma_i}^{\emptyset} = 0) \left[ u(|t_R - t_i|) - \theta \left( \frac{1}{2} + \frac{1}{2n} \right) \right] + \sum_{j=1}^{e-1} p(X_{\sigma_i}^{\emptyset} = j) \left[ u(|t_{RE}^s - t_i|) - \bar{d} - \theta \left( \frac{1}{2} + \frac{1}{2n} - \frac{j}{n} \right) \right] + p(X_{\sigma_i}^{\emptyset} = e) \left[ u(|t_{RE}^s - t_i|) - \bar{d} - \theta \left( \frac{1}{2} (1 - s_E) + \frac{1}{2n} \right) \right] + \sum_{j=e+1}^{2e-1} p(X_{\sigma_i}^{\emptyset} = j) \left[ u(|t_{LE}^s - t_i|) - \bar{d} - \theta \left( \frac{1}{2} (1 - s_E) - \frac{1}{2n} + \frac{j - e}{n} \right) \right] + p(X_{\sigma_i}^{\emptyset} = 2e) \left[ u(|t_L - t_i|) - \theta \left( \frac{1}{2} + \frac{1}{2n} \right) \right].
$$
Therefore,

\[
\mathbb{E} \left[ U_i((L, \sigma_{-i})) \right] - \mathbb{E} \left[ U_i((R, \sigma_{-i})) \right] = \frac{1}{n} \sum_{j=0}^{2e-1} p(X^\sigma_{-i} = j) \left[ u(|t^w_{RE} - t_i|) - u(|t_R - t_i|) - \bar{d} - \theta \left( \frac{1}{2} - \frac{1}{2n} \right) \right]
+ \sum_{j=1}^{2e} p(X^\sigma_{-i} = j) \left[ u(|t^w_{RE} - t_i|) - \bar{d} - \theta \left( \frac{1}{2} - \frac{1}{2n} - \frac{j}{n} \right) \right]
+ p(X^\sigma_{-i} = 2e) \left[ u(|t_L - t_i|) - \theta \left( \frac{1}{2} + \frac{1}{2n} \right) \right] .
\]

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

On the one hand,

\[
\mathbb{E} \left[ U_i((L, \sigma_{-i})) \right] = p(X^\sigma_{-i} = 0) \left[ u(|t^w_{RE} - t_i|) - \bar{d} - \theta \left( \frac{1}{2} - \frac{1}{2n} \right) \right]
+ \sum_{j=1}^{2e-1} p(X^\sigma_{-i} = j) \left[ u(|t^w_{RE} - t_i|) - \bar{d} - \theta \left( \frac{1}{2} - \frac{1}{2n} - \frac{j}{n} \right) \right]
+ p(X^\sigma_{-i} = 2e) \left[ u(|t_L - t_i|) - \theta \left( \frac{1}{2} + \frac{1}{2n} \right) \right] .
\]

On the other hand,

\[
\mathbb{E} \left[ U_i((R, \sigma_{-i})) \right] = p(X^\sigma_{-i} = 0) \left[ u(|t_R - t_i|) - \theta \left( \frac{1}{2} + \frac{1}{2n} \right) \right]
+ \sum_{j=1}^{2e-1} p(X^\sigma_{-i} = j) \left[ u(|t^w_{RE} - t_i|) - \bar{d} - \theta \left( \frac{1}{2} + \frac{1}{2n} - \frac{j}{n} \right) \right]
+ p(X^\sigma_{-i} = 2e) \left[ u(|t^w_{RE} - t_i|) - \bar{d} - \theta \left( \frac{1}{2} - s_E + \frac{1}{2n} \right) \right] .
\]

Therefore,

\[
\mathbb{E} \left[ U_i((L, \sigma_{-i})) \right] - \mathbb{E} \left[ U_i((R, \sigma_{-i})) \right] = p(X^\sigma_{-i} = 0) \left[ u(|t^w_{RE} - t_i|) - u(|t_R - t_i|) - \bar{d} + \frac{1}{n} \theta \right]
+ \sum_{j=1}^{2e-1} p(X^\sigma_{-i} = j) \frac{1}{n} \theta
+ p(X^\sigma_{-i} = 2e) \left[ u(|t_L - t_i|) - u(|t^w_{RE} - t_i|) + \bar{d} - s_E \theta \right]
= p(X^\sigma_{-i} = 0) \left[ u(|t^w_{RE} - t_i|) - u(|t_R - t_i|) - \bar{d} \right]
+ p(X^\sigma_{-i} = 2e) \left[ u(|t_L - t_i|) - u(|t^w_{RE} - t_i|) + \bar{d} - \left( s_E + \frac{1}{n} \right) \theta \right] + \frac{1}{n} \theta .
\]

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

Therefore,

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

On the other hand,

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

Therefore,

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

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

On the one hand,

\[ \mathbb{E} \left[ U_i((L, \sigma_{-i})) \right] \]
\[ = p(\mathcal{X}_i^\sigma = 0) \left[ u \left( \left| \left( \frac{1}{2} - s_E + \frac{1}{2n}, \frac{1}{2} - \frac{1}{2n} \right) - t_i \right| \right) - \theta(1 - s_E) \right] \]
\[ + \sum_{j=1}^{2e-1} p(\mathcal{X}_i^{\sigma} = j) \left[ u \left( \left| \left( \frac{1}{2} - s_E + \frac{j}{n}, \frac{1}{2} - \frac{j}{n} \right) - t_i \right| \right) - \theta(1 - s_E) \right] \]
\[ + p(\mathcal{X}_i^{\sigma} = 2e) \left[ u(|t_L - t_i|) - \theta \left( \frac{1}{2} + \frac{1}{2n} \right) \right]. \]
On the other hand,

\[
\mathbb{E} \left[ U_i((R, \sigma_{-i})) \right] \\
= p(\chi^\sigma_{-i} = 0) \left[ u \left( |t_R - t_i| - \theta \left( \frac{1}{2} + \frac{1}{2n} \right) \right) \right] \\
+ \sum_{j=1}^{2e-1} p(\chi^\sigma_{-i} = j) \left[ u \left( \left| t \left( \frac{1}{2} - s_{E} - \frac{1}{2n} + \frac{j}{n\cdot2} + \frac{1}{2n} - \frac{j}{n} \right) - t_i \right| \right) - \theta \left( 1 - s_{E} \right) \right] \\
+ p(\chi^\sigma_{-i} = 2e) \left[ u \left( \left| t \left( \frac{1}{2} - \frac{1}{2n} - s_{E} + \frac{1}{2n} \right) - t_i \right| \right) - \theta \left( 1 - s_{E} \right) \right].
\]

Therefore,

\[
\mathbb{E} \left[ U_i((L, \sigma_{-i})) \right] - \mathbb{E} \left[ U_i((R, \sigma_{-i})) \right] \\
= p(\chi^\sigma_{-i} = 0) \left[ u \left( \left| t \left( \gamma + \frac{1}{2n} - \frac{1}{2n} \right) - t_i \right| \right) - u(|t_R - t_i|) - \theta \left( \gamma - \frac{1}{2n} \right) \right] \\
+ \sum_{j=1}^{2e-1} p(\chi^\sigma_{-i} = j) \left[ u \left( \left| t \left( \gamma + \frac{1}{2n} + \frac{j}{n\cdot2} + \frac{1}{2n} - \frac{j}{n} \right) - t_i \right| \right) \\
- u \left( \left| t \left( \gamma - \frac{1}{2n} + \frac{j}{n\cdot2} + \frac{1}{2n} - \frac{j}{n} \right) - t_i \right| \right) \right] \\
+ p(\chi^\sigma_{-i} = 2e) \left[ u(|t_L - t_i|) - u \left( \left| t \left( \frac{1}{2} - \frac{1}{2n} + \frac{1}{2n} \right) - t_i \right| \right) + \theta \left( \gamma - \frac{1}{2n} \right) \right].
\]
Proof of Lemma 2

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

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

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

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

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

$$
E[1_{I=I}] = P(I = I) = \left( \frac{|N| - 1}{|I|} \right) |I|^{|N|-|I|-1},
$$

and so the above can be written as

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

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

59
Proof of Lemma 3

(iii) ⇔ (iv).

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

(i) ⇒ (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 (16)) and equation (18), we have

\[ \phi_i^C(\sigma_i'', \sigma_i', \sigma_i') \geq 0 \quad \text{and} \quad \phi_i^C(\sigma_i'', \sigma_i', \sigma_i') \leq 0, \]

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

\[ \varepsilon = \varepsilon^0 := \frac{\phi_i^C(\sigma_i'', \sigma_i', \sigma_i') - \phi_i^C(\sigma_i'', \sigma_i', \sigma_i')}{\phi_i^C(\sigma_i'', \sigma_i', \sigma_i') + \phi_i^C(\sigma_i'', \sigma_i', \sigma_i')}, \]

and thus the maximum level at which \( \sigma'' \) is immune to correlated errors w.r.t. \( \sigma' \) for player \( i \) is precisely \( \varepsilon^0 \). In turn, using antisymmetry (equation (18)), this implies that

\[ \sigma' \succsim_i \sigma'' \quad \text{(respectively} \succsim_i \text{for the strict domination case)} \]  

\[ \Leftrightarrow \forall \varepsilon \leq \varepsilon^0 (1 - \varepsilon)\phi_i^C(\sigma_i', \sigma_i', \sigma_i') - \varepsilon\phi_i^C(\sigma_i'', \sigma_i', \sigma_i') \geq 0 \quad \text{(respectively} > 0 \text{)} \]  

(29)

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 (29) must be < 0 (respectively \( \leq 0 \)). Contradiction to weak (respectively strict) domination \( \sigma' \succsim_i \sigma'' \).

(iii) ⇒ (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

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

so that inequality (29) (respectively the strict version) holds.

(iii) ⇒ (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 $\sigma'$ being immune to correlated errors w.r.t. $\sigma''$. $h_i$ is almost identical to $g_i$, except that it has its unique 0 at $1 - \varepsilon^0$, rather than $\varepsilon^0$. 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 $\delta$. 

61
Proof of Lemma 4

Let $f(\varepsilon) := f_i(\varepsilon)$ be defined as in the statement of the Lemma, where for notational convenience we now drop the subscript $i$. The key to this Lemma is recognizing that by antisymmetry of $\phi$ in its first two arguments (equation (18)), we have

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

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

(i).

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

$$f(\varepsilon) \geq 0 \quad \forall \varepsilon \in \left[0, \frac{1}{2}\right]. \quad (30)$$

But this immediately implies that $-f(1 - \varepsilon) \leq 0 \quad \forall \varepsilon \in \left[\frac{1}{2}, 1\right]$. By the remark above, this already proves $\sigma' \succ^+ \sigma''$, provided that we show that the pathological case, where $\exists \varepsilon^* > \frac{1}{2}$ such that $-f(1 - \varepsilon) = 0 \quad \forall \varepsilon \in \left[\frac{1}{2}, \varepsilon^*\right]$ while $f(\varepsilon) \leq 0 \quad \forall \varepsilon \in \left[\frac{1}{2}, 1\right]$ and $\exists \varepsilon \in \left[\frac{1}{2}, 1\right]$ such that $f(\varepsilon) < 0$ (hence $-f(1 - \varepsilon) \geq 0 \quad \forall \varepsilon \in \left[0, \frac{1}{2}\right]$), cannot occur. In that case, $\sigma''$ would be immune to uncorrelated errors w.r.t. $\sigma'$ at level $\varepsilon^*$, 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 $\left[\frac{1}{2} - \delta, \frac{1}{2}\right]$ $(\delta > 0)$ of $\frac{1}{2}$ must contain a point $\varepsilon$ where $f(\varepsilon) > 0$ (recall that (30) holds). But then also every right neighbourhood $\left[\frac{1}{2}, \frac{1}{2} + \delta\right]$ must contain a point $\varepsilon$ where $-f(1 - \varepsilon) < 0$, and thus $\sigma''$ can be immune to uncorrelated at level at $\frac{1}{2}$ w.r.t. $\sigma'$ 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 $\sigma' \sim^+ \sigma''$.

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

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

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

For the strict case, note that with our “near-monotonicity” property, if \( \sigma' \succ_i \sigma'' \), \( f \left( \frac{1}{2} \right) = 0 \) is no longer possible, as this would imply

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

from which it would follow that both \( \sigma' \) and \( \sigma'' \) are immune w.r.t. each other at maximal level exactly \( \frac{1}{2} \) and so \( \sigma'_i \sim_i \sigma''_i \). As noted above, by continuity, \( f(\varepsilon) \geq 0 \ \forall \varepsilon \in \left[ 0, \frac{1}{2} \right] \) and \( f \left( \frac{1}{2} \right) > 0 \) is equivalent to immunity of \( \sigma' \) w.r.t. \( \sigma'' \) at level \( \frac{1}{2} + \delta \) for some \( \delta > 0 \).
Proof of Lemma 5

Let $f(\varepsilon) := f_i(\varepsilon)$, as in the statement of the Lemma, where for convenience we drop the subscript $i$. Analogously to the proof of Lemma 4, we use the fact that

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

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

For the strict domination case, let $\varepsilon^0_{\sigma'', \sigma'}$ be the maximum level at which $\sigma''$ is immune to uncorrelated errors w.r.t. $\sigma'$ for $i$. In other words, $\varepsilon^0_{\sigma'', \sigma'}$ is the smallest root of $-f(1 - \varepsilon)$ in $[0, 1]$. By assumption, we must have $f(\varepsilon^0_{\sigma'', \sigma'}) > -f(1 - \varepsilon^0_{\sigma'', \sigma'}) = 0$. But then by continuity and the fact that $f(\varepsilon) > -f(1 - \varepsilon) \, \forall \varepsilon \in [0, 1]$, the smallest root of $f$ in $[0, 1]$, call it $\varepsilon^0_{\sigma', \sigma''}$, must be strictly greater than $\varepsilon^0_{\sigma'', \sigma'}$. This means that $\sigma'$ is immune at level $\varepsilon^0_{\sigma', \sigma''}$, w.r.t. $\sigma''$, but not the other way around, so that $\sigma'' \not\leq^i \sigma'$.
Proof of Lemma 6

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

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

and after rewriting

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

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 Lemma 5 shows \( \sigma^R \succ_i \sigma^L \ \forall i \in \Omega^R \). (Recall that \( \phi_i^C (\sigma_i^R, \sigma_i^L, (1 - \varepsilon)\sigma_{-i}^R + \varepsilon \sigma_{-i}^L) = -f_i(1 - \varepsilon) \).) The key to this is to show:

\[\sum_{j=0}^{\varepsilon} p(B_{1-\varepsilon}^{2e} = j) = \sum_{j=\varepsilon+1}^{2e} p(B_{1-\varepsilon}^{2e} = \bar{j}).\]

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

\[\binom{2e}{j} (1 - \varepsilon)^j \varepsilon^{2e-j},\]

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

With this result, we now have

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

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:

\[
K_2 + K_3 = u(|t_{RE}^i - t_i|) - u(|t_R - t_i|) + u(t_i - t_L) - u(t_i - t_{LE}^i).
\]

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 \( t_i \in \left( \frac{1}{2}, t_R \right) \). For \( t_i \in \left( \frac{1}{2}, t_{RE}^i \right) \):

\[
\frac{\partial}{\partial t_i} \left\{ u(|t_{RE}^i - t_i|) - u(|t_R - t_i|) + u(t_i - t_L) - u(t_i - t_{LE}^i) \right\} \\
= -u'(t_i - t_{RE}^i) + u'(t_R - t_i) + u'(t_i - t_L) - u'(t_i - t_{LE}^i) \leq 0,
\]

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

\[
\frac{\partial}{\partial t_i} \left\{ u(|t_{RE}^i - t_i|) - u(|t_R - t_i|) + u(t_i - t_L) - u(t_i - t_{LE}^i) \right\} \\
= u'(t_i - t_{RE}^i) + u'(t_R - t_i) + u'(t_i - t_L) - u'(t_i - t_{LE}^i) \leq 0,
\]

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 \( f_i(\varepsilon) + f_i(1 - \varepsilon) < 0 \).

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 \( K_3, \sigma^R \) is not immune to uncorrelated errors at level \( \frac{1}{2} \) w.r.t. \( \sigma^L \). To illustrate these examples, and also coincidentally add intuition about our \( \phi^C_i \) function in the case of uncorrelated errors, Figures 1 to 6 show the graphs of \( f_i(\varepsilon) = \phi^C_i(\sigma^L_i, \sigma^R_i, (1 - \varepsilon)\sigma_{-i} + \varepsilon\sigma_{-i}^R) \) and \( -f_i(1 - \varepsilon) = \phi^C_i(\sigma^R_i, \sigma^L_i, (1 - \varepsilon)\sigma_{-i}^R + \varepsilon\sigma_{-i}^L) \) for different choices of the (model dependent) vector of parameters \( (n, e, K_1, K_2, K_3) \), while \( \theta \) 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 \( u(x) = -\frac{1}{2}x^2, t_L = 0.2, t_R = 0.8, t_{LE} = 0.4, t_{RE} = 0.6, t_i = 0.65, \) and \( d = 0.1 \). 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 \( \phi^C_i(\sigma^R_i, \sigma^L_i, (1 - \varepsilon)\sigma_{-i}^R + \varepsilon\sigma_{-i}^L) \)
Figure 1: Plot of $\phi(\cdot)$ for $(n, e, K_1, K_2, K_3) = (21, 3, -1, -0.5, 0.3)$ occurs before 0.2. That is, $\sigma^R$ is immune to uncorrelated errors w.r.t. $\sigma^L$ 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 $\sigma^R$ can be immune to uncorrelated errors w.r.t. $\sigma^L$ at most at a level smaller than 0.2.
Figure 2: Plot of $\phi(\cdot)$ for $(n, e, K_1, K_2, K_3) = (41, 6, -1, -0.5, 0.3)$
Figure 3: Plot of $\phi(\cdot)$ for $(n, e, K_1, K_2, K_3) = (81, 12, -1, -0.5, 0.3)$
Figure 4: Plot of $\phi(\cdot)$ for $(n, e, K_1, K_2, K_3) = (21, 3, -0.03, -0.09, 0.03)$
Figure 5: Plot of $\phi(\cdot)$ for $(n, e, K_1, K_2, K_3) = (41, 6, -0.03, -0.09, 0.03)$
Figure 6: Plot of $\phi(\cdot)$ for $(n, e, K_1, K_2, K_3) = (81, 12, -0.03, -0.09, 0.03)$
Proof of Proposition 4

(i) and (iii).

We prove that $\sigma^R \preceq \sigma^{LR}$ (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 $\chi_{-i}^R$ and $\chi_{-i}^{LR}$, in order to use Lemma 1 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

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

and

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

Thus according to Lemma 1,

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

while

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

and hence

$$\phi_i^C(\sigma_i^R, \sigma_i^{LR}, \sigma_i^R) - \phi_i^C(\sigma_i^{LR}, \sigma_i^R, \sigma_i^{LR}) = -\frac{3\theta}{2n} + \bar{d} + u(t_R - t_i) - \frac{1}{2} \{u(|t_{LE} - t_i|) + u(t_{RE} - t_i)\}$$

$$= \frac{1}{2} \left[ \bar{d} + u(t_R - t_i) - u(|t_{LE} - t_i|) - \frac{3\theta}{n} \right]$$

$$+ \frac{1}{2} \left[ \bar{d} + u(t_R - t_i) - u(t_{RE} - t_i) - \frac{3\theta}{n} \right] > 0$$

where each of the terms in brackets is strictly greater than 0 by Corollary 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
\[
\phi^C_i (\sigma^R_m, \sigma^L_m, \sigma^R_{-m}) = \frac{1}{2} \left\{ u(t_R - t_m) - u(t^*_R - t_m) + \bar{d} - \frac{\theta}{n} \right\},
\]
\[
\phi^C_i (\sigma^L_m, \sigma^R_m, \sigma^L_{-m}) = 0.
\]

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
\[
E \left[ U_i (L, \sigma^L_{-m}) \mid (\varnothing, \varnothing) \right] = E \left[ U_i (R, \sigma^R_{-m}) \mid (\varnothing, \varnothing) \right].
\]

However, the expression for \( \phi^C_i (\sigma^R_m, \sigma^L_m, \sigma^R_{-m}) \) is strictly greater than 0, by Corollary 1.(iii) and the fact that \( e \geq 1 \) is always assumed.

This shows \( \sigma^R \succ \sigma^L \) in the presence of correlated errors. As already remarked, by Lemma 3, this also shows that \( \sigma^R \) dominates \( \sigma^L \) according to maximum entropy in equilibria. The proof for \( \sigma^L \succ \sigma^L \) 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.

\( \text{(iv).} \)

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

Consider first \( i \in \Omega^L \).

The distribution of the number of left-wing votes \( \chi^\frac{1}{2} \sigma^L_i + \frac{1}{2} \sigma^R_i \) is given by:
\[
p(\chi^\frac{1}{2} \sigma^L_i + \frac{1}{2} \sigma^R_i = j) = \begin{cases} 
(\frac{1}{2})^{e-1} (\frac{3}{4}) & \text{if } j = 0, \\
(\frac{1}{2})^{e-1} \left( \frac{1}{4} (j-1) + \frac{3}{4} (e-1) \right) & \text{if } 1 \leq j \leq e - 1, \\
(\frac{1}{2})^{e-1} (\frac{1}{4}) & \text{if } j = e, \\
0 & \text{if } e + 1 \leq j \leq 2e.
\end{cases}
\]
Then, by Lemma 1:

\[
\phi_i^C \left( \sigma_{LR}^i, \sigma_{i}^R, \frac{1}{2} \sigma_{-i}^{LR} + \frac{1}{2} \sigma_{-i}^R \right) = \phi_i^C \left( L, R, \frac{1}{2} \sigma_{-i}^{LR} + \frac{1}{2} \sigma_{-i}^R \right) 
\]

\[
= \frac{\theta}{n} \sum_{j=0}^{e-1} p(X_{-i} = j) + p(X_{-i} = e) [u(t_{LE}^i - t_i)] - u(t_{RE}^i - t_i)] 
+ p(X_{-i} = 0) [u(t_{RE}^i - t_i) - u(t^R - t_i) - d] 
\]

\[
= \left( 1 - \frac{1}{4} \left( \frac{1}{2} \right) ^{(e-1)} \right) \frac{\theta}{n} \left[ u(t_{LE}^i - t_i) - u(t_{RE}^i - t_i) \right] 
+ \frac{3}{4} \left( \frac{1}{2} \right) ^{(e-1)} [u(t_{RE}^i - t_i) - u(t^R - t_i) - d]. 
\]

Let

\[
K_1 = u(t_{LE}^i - t_i) - u(t_{RE}^i - t_i) 
K_2 = u(t_{RE}^i - t_i) - u(t^R - t_i) - d
\]

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

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

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

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

75
\[ K_2 \quad N = \text{minimal odd } n \quad e = \text{rd} \left( \frac{\tilde{s}_E N}{2} \right) \quad \phi(\cdot) \text{ at } N \]

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.5</td>
<td>45</td>
<td>6</td>
<td>1.03e-2</td>
<td></td>
</tr>
<tr>
<td>-1</td>
<td>53</td>
<td>7</td>
<td>3.98e-3</td>
<td></td>
</tr>
<tr>
<td>-3</td>
<td>69</td>
<td>9</td>
<td>5.69e-3</td>
<td></td>
</tr>
<tr>
<td>-10</td>
<td>85</td>
<td>11</td>
<td>4.44e-3</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Example calculations for \( \phi^C_i \left( \sigma_{i LR}^i, \sigma_{i R}^i, \frac{1}{2} \sigma_{-i}^{LR} + \frac{1}{2} \sigma_{-i}^R \right) \) 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 \( \phi(\cdot) > 0. \)

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

Table 4: Example calculations for \( \phi^C_i \left( \sigma_{i LR}^i, \sigma_{i R}^i, \frac{1}{2} \sigma_{-i}^{LR} + \frac{1}{2} \sigma_{-i}^R \right) \) 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 \( \phi(\cdot) > 0. \)

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

\[
K_2 \geq K_2 \left( t_i = \frac{1}{2} \right) = u \left( t_{RE}^i - \frac{1}{2} \right) - u \left( t_R - \frac{1}{2} \right) - \tilde{d},
\]

so that for fixed choices of \( t_R, t_{RE}^i, \tilde{d}, u(\cdot), \) and \( \tilde{s}_E, \) a large enough \( N \in \mathbb{N} \) such that \( \phi^C_i \left( \sigma_{i LR}^i, \sigma_{i R}^i, \frac{1}{2} \sigma_{-i}^{LR} + \frac{1}{2} \sigma_{-i}^R \right) > 0 \) can be chosen uniformly over the player set \( \Omega^L \) (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 \( \phi^C_i \left( \sigma_{i LR}^i, \sigma_{i R}^i, \frac{1}{2} \sigma_{-i}^{LR} + \frac{1}{2} \sigma_{-i}^R \right) < 0, \) 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 \( \phi^C_i \left( \sigma_{i LR}^i, \sigma_{i R}^i, \frac{1}{2} \sigma_{-i}^{LR} + \frac{1}{2} \sigma_{-i}^R \right) \) as a function of \( n \) for some possible parameter values. \( \theta \) 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, \( \phi \) 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 $\phi$ are always quite small, so that the domination according to maximum entropy in strategies (version 1) is not very robust.

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

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

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

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

$$\phi^{C}_{m} \left( \sigma_{m}^{LR}, \sigma_{m}^{R}, \frac{1}{2} \sigma_{-m}^{LR} + \frac{1}{2} \sigma_{-m}^{R} \right)$$

$$= \frac{1}{2} \mathbb{E} \left[ U_{m} \left( L, \frac{1}{2} \sigma_{-m}^{LR} + \frac{1}{2} \sigma_{-m}^{R} \right) \right] + \frac{1}{2} \mathbb{E} \left[ U_{m} \left( R, \frac{1}{2} \sigma_{-m}^{LR} + \frac{1}{2} \sigma_{-m}^{R} \right) \right]$$

$$- \mathbb{E} \left[ U_{m} \left( R, \sigma_{m}^{LR} \right) \right]$$

$$= \frac{1}{2} \phi^{C}_{m} \left( L, R, \frac{1}{2} \sigma_{-m}^{LR} + \frac{1}{2} \sigma_{-m}^{R} \right).$$

Thus by Lemma 1:

$$2\phi^{C}_{m} \left( \sigma_{m}^{LR}, \sigma_{m}^{R}, \frac{1}{2} \sigma_{-m}^{LR} + \frac{1}{2} \sigma_{-m}^{R} \right)$$

$$= \left(1 - \left(\frac{1}{2}\right)^{e}\right) \frac{\theta}{n} + \left(\frac{1}{2}\right)^{e} \left[ u \left( t_{RE}^{s} - \frac{1}{2} \right) - u \left( t_{R} - \frac{1}{2} \right) - \bar{d} \right].$$

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 $\sigma^{LR} > \sigma^{R}$ according to the maximum entropy in strategies (version 1) criterion.

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

$$\phi^{C}_{i} \left( \sigma_{i}^{LR}, \sigma_{i}^{R}, \frac{1}{2} \sigma_{-i}^{LR} + \frac{1}{2} \sigma_{-i}^{R} \right)$$

$$= \left(1 - \frac{1}{4} \left(\frac{1}{2}\right)^{e-1}\right) \frac{\theta}{n} - \frac{3}{4} \left(\frac{1}{2}\right)^{e-1} \left[ u \left( t_{i} - t_{LE}^{s} \right) - u \left( t_{L} \right) \right]$$

$$- \frac{3}{4} \left(\frac{1}{2}\right)^{e-1} \left[ u \left( t_{i} - t_{L} \right) - u \left( t_{LE}^{s} - t_{i} \right) + \bar{d} \right].$$

77
while for \( i = m, \chi_{-m}^{\frac{1}{2} \sigma_{LR} + \frac{1}{2} \sigma_L} = e + B_{1/2}^{\epsilon} \), so that

\[
2 \phi_m^C (\sigma_m^{LR} : \sigma_m^L, \frac{1}{2} \sigma_{LR} + \frac{1}{2} \sigma_m^L) \\
= \mathbb{E} \left[ U_m \left( R, \frac{1}{2} \sigma_{LR} + \frac{1}{2} \sigma_m^L \right) \right] - \mathbb{E} \left[ U_m \left( L, \frac{1}{2} \sigma_{LR} + \frac{1}{2} \sigma_m^R \right) \right] \\
= \left( 1 - \left( \frac{1}{2} \right)^e \right) \frac{\theta}{n} - \left( \frac{1}{2} \right)^e \left[ u \left( \frac{1}{2} - t_L \right) - u \left( \frac{1}{2} - t_{LE}^i \right) + d \right].
\]

(ii).

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

\[
f_i(\epsilon) = \phi_i^C (\sigma_i^{LR}, \sigma_i^L, (1 - \epsilon)\sigma_{-i}^{LR} + \epsilon \sigma_{-i}^L)
\]

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

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

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

and there is nothing more to show.

Thus consider \( i \in \Omega^R \).

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

\[
p \left( \chi_{-i}^{(1-\epsilon)\sigma_{LR} + \epsilon \sigma_L} = j \right) = \begin{cases} 
0 & \text{if } j < e, \\
\frac{1 - \epsilon}{2} p \left( B_{\epsilon}^{e-1} = 0 \right) & \text{if } j = e, \\
\frac{1 + \epsilon}{2} p \left( B_{\epsilon}^{e-1} = j - e - 1 \right) + \frac{1 + \epsilon}{2} p \left( B_{\epsilon}^{e-1} = j - e \right) & \text{if } e + 1 \leq j \leq 2e - 1, \\
\frac{1 + \epsilon}{2} p \left( B_{\epsilon}^{e-1} = e - 1 \right) & \text{if } j = 2e.
\end{cases}
\]
Using Lemma 1, we calculate:

\[
\phi_i^C (\sigma^2_{i-1}, \sigma^L_{i-1}, (1 - \varepsilon)\sigma^L_{i-1} + \varepsilon\sigma^L_{i-1}) = -\phi_i^C (\sigma^2_{i-1}, \sigma^L_{i-1}, (1 - \varepsilon)\sigma^L_{i-1} + \varepsilon\sigma^L_{i-1})
\]
\[
= \frac{\theta}{n} \left( 1 - \left( \frac{1 - \varepsilon}{2} \right)^2 \right) - \frac{(1 - \varepsilon)^2}{2} \left[ u(t_i - t^*_{RE}) - u(t_i - t^*_{LE}) \right]
\]
\[
- \left( \frac{1 + \varepsilon}{2} \right) \varepsilon e^{-1} \left[ u(t_i - t_L) - u(t_i - t^*_{LE}) + d \right]
\]
\[
= \frac{\theta}{n} \left( 1 - \left( \frac{1 - \varepsilon}{2} \right)^2 \right) + (1 - \varepsilon)^e \frac{K_1}{2} + (1 + \varepsilon)^e e^{-1} \frac{K_2}{2},
\]

where this time we define

\[
K_1 := u \left( |t_i - t^*_{RE}| \right) - u(t_i - t^*_{LE}),
\]
\[
K_2 := u(t_i - t^*_{LE}) - u(t_i - t_L) - d.
\]

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 \(\varepsilon\), while the two other terms are decreasing. Differentiating w.r.t. \(\varepsilon\) yields:

\[
\frac{\partial}{\partial \varepsilon} \phi_i^C (\sigma^2_{i-1}, \sigma^L_{i-1}, (1 - \varepsilon)\sigma^L_{i-1} + \varepsilon\sigma^L_{i-1})
\]
\[
= \frac{\theta}{n} \frac{e}{2} (1 - \varepsilon)^{e-1} - \frac{eK_1}{2} (1 - \varepsilon)^{e-1} + \frac{(2e - 1)K_2}{2} (1 + \varepsilon)^e e^{-2}
\]
\[
= \frac{e}{2} (1 - \varepsilon)^e \left( \frac{\theta}{n} - K_1 \right) + \frac{(2e - 1)K_2}{2} (1 + \varepsilon)^e e^{-2}
\]

Suppose that \(K_1 > 0\) is fixed. Then for large enough \(n\), the above derivative will be negative, and \(\phi_i^C (\sigma^2_{i-1}, \sigma^L_{i-1}, (1 - \varepsilon)\sigma^L_{i-1} + \varepsilon\sigma^L_{i-1})\) 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

\[
\frac{\partial}{\partial \varepsilon} \phi_i^C (\sigma^2_{i-1}, \sigma^L_{i-1}, (1 - \varepsilon)\sigma^L_{i-1} + \varepsilon\sigma^L_{i-1}) < 0 \quad \forall n \geq N \forall i \in \Omega^R(n).
\]

In order to be able to make statements about the derivative above, and hence the monotonicity of \(f_i(\varepsilon) = \phi_i^C (\sigma^2_{i-1}, \sigma^L_{i-1}, (1 - \varepsilon)\sigma^L_{i-1} + \varepsilon\sigma^L_{i-1})\) w.r.t. \(\varepsilon\), uniformly for a subset of players in \(\Omega^R\), 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 \( \phi^C_i (\sigma_{i}^{LR}, \sigma_i^L, (1 - \varepsilon)\sigma_{-i}^{LR} + \varepsilon\sigma_{-i}^L) \geq 0 \ \forall \varepsilon \in [0, \frac{1}{2}] \), that is, that \( \sigma^{LR} \) is immune to uncorrelated errors at level \( \frac{1}{2} \) w.r.t. \( \sigma^L \) for any player \( i \in \Omega^R \).

By Lemma 4.(i), this implies that \( \sigma^{LR} \succ_i \sigma^L \ \forall i \in \Omega^R \). Consider

\[
\frac{\partial^2}{\partial \varepsilon^2} \phi^C_i (\sigma_{i}^{LR}, \sigma_i^L, (1 - \varepsilon)\sigma_{-i}^{LR} + \varepsilon\sigma_{-i}^L) = \frac{-(e - 1)e}{2} (1 - \varepsilon)^{e-2} \left( \frac{\theta}{n} - K_1 \right) + \frac{(2e - 1)(2e - 3)K_2}{2} (1 + \varepsilon)^{e-3}
\]

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

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

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

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

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 \( \varepsilon^* = \frac{1}{2} \ \forall i \in \Omega^R \).

Now consider the median voter \( i = m \).

As above, we show that

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

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

\[
2\phi^C_m (\sigma_{m}^{LR}, \sigma_m^L, (1 - \varepsilon)\sigma_{-m}^{LR} + \varepsilon\sigma_{-m}^L) = -\phi^C_i (L, R, (1 - \varepsilon)\sigma_{-m}^{LR} + \varepsilon\sigma_{-m}^L)
\]

\[
= (1 - (1 - \varepsilon)e) \frac{\theta}{n} + \varepsilon e \left[ u \left( \frac{1}{2} - t_{LE}^* \right) - u \left( \frac{1}{2} - t_L \right) - \tilde{d} \right]
\]

\[
= (1 - (1 - \varepsilon)e) \frac{\theta}{n} + \varepsilon e K.
\]

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

\[
\frac{\partial}{\partial \varepsilon} f_m(\varepsilon) = e \left\{ \frac{\theta}{n} (1 - \varepsilon)^{e-1} + \varepsilon^{e-1} K \right\} < 0
\]

\[
\iff \frac{\theta}{nK} < \left( \frac{\varepsilon}{1 - \varepsilon} \right)^{e-1}
\]

80
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 $\phi^C_i(\sigma^{LR}_i, \sigma^R_i, (1 - \varepsilon)\sigma^{LR}_{-i} + \varepsilon\sigma^R_{-i})$.

(v) and (vii).

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

$$\phi^C_i(\sigma^L, \sigma^R, \sigma^L_{-i}) - \phi^C_i(\sigma^R, \sigma^L, \sigma^R_{-i})$$

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

$$\phi^C_i(\sigma^L, \sigma^R, \sigma^L_{-i}) = u(|t_i - t_L|) - \theta \left( \frac{1}{2} + \frac{1}{2n} \right) - u(|t_{LE}^i - t_i|) + \theta \left( \frac{1}{2} - \frac{1}{2n} \right) + \bar{d},$$

$$\phi^C_i(\sigma^R, \sigma^L, \sigma^R_{-i}) = u(|t_i - t_R|) - \theta \left( \frac{1}{2} + \frac{1}{2n} \right) - u(|t_{RE}^i - t_i|) + \bar{d} + \theta \left( \frac{1}{2} - \frac{1}{2n} \right).$$

Thus

$$\phi^C_i(\sigma^L, \sigma^R, \sigma^L_{-i}) - \phi^C_i(\sigma^R, \sigma^L, \sigma^R_{-i}) = u(|t_i - t_L|) - u(|t_i - t_{LE}^i|) + u(|t_i - t_{RE}^i|) - u(|t_i - t_R|).$$

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 $t_i > t_m$.

(vi).

Lemma 6 already established that $\sigma^R \succ \perp \sigma^L \forall i \in \Omega^R$. We take the results derived in Lemma 6:

$$f_i(\varepsilon) := \phi^C_i(\sigma^L_i, \sigma^R_i, (1 - \varepsilon)\sigma^{LR}_{-i} + \varepsilon\sigma^R_{-i})$$

$$= 2p(B^{2e}_{1-\varepsilon} = e)K_1 + \{\varepsilon^{2e} + (1 - \varepsilon)^{2e}\} [K_2 + K_3],$$
where recall that
\[
K_1 := u(|t_i - t_{LE}^*|) - u(|t_{RE}^* - t_i|),
\]
\[
K_2 := u(|t_{RE}^* - t_i|) - u(|t_R - t_i|) - \bar{d},
\]
\[
K_3 := u(|t_i - t_L^*|) - u(|t_i - t_{LE}^*|) + \bar{d},
\]
and show that \(f_i(\varepsilon) + f_i(1 - \varepsilon) > 0 \forall i \in \Omega^L\), which by Lemma 5, proves that \(\sigma^L \succ_i \sigma^R \forall i \in \Omega^L\). 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.

(viii).

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 \(i \in \Omega^R\).

We have \(\chi_{-i}^{\frac{1}{2}\sigma^L + \frac{1}{2}\sigma^R} \sim \text{Bin}\left(\frac{1}{2}, 2e\right)\). Thus:
\[
\phi_i^C\left(\sigma^L, \sigma^R, \frac{1}{2}\sigma_{-i}^L + \frac{1}{2}\sigma_{-i}^R\right) = 
\sum_{j=1}^{e-1} p(\chi_{-i}^{\frac{1}{2}\sigma^L + \frac{1}{2}\sigma^R} = j) \left[ u(|t_{RE}^* - t_i|) - \bar{d} - \theta \left(\frac{1}{2} + \frac{1}{2n} - \frac{j}{n} + \frac{1}{n}\right) 
- u(|t_{RE}^* - t_i|) + \bar{d} + \theta \left(\frac{1}{2} + \frac{1}{2n} - \frac{j}{n} + \frac{1}{n}\right) \right] 
+ \sum_{j=e+1}^{2e-1} p(\chi_{-i}^{\frac{1}{2}\sigma^L + \frac{1}{2}\sigma^R} = j) \left[ -2\frac{\theta}{n} \right] 
+ p(\chi_{-i}^{\frac{1}{2}\sigma^L + \frac{1}{2}\sigma^R} = 0) \left[ u(|t_{RE}^* - t_i|) - \bar{d} - \theta \left(\frac{1}{2} - \frac{1}{2n}\right) - u(|t_R - t_i|) + \theta \left(\frac{1}{2} + \frac{1}{2n}\right) \right] 
+ p(\chi_{-i}^{\frac{1}{2}\sigma^L + \frac{1}{2}\sigma^R} = e) \left[ u(|t_{LE}^* - t_i|) 
- \bar{d} - \theta \left(\frac{1}{2} + \frac{1}{2n}\right) - u(|t_{LE}^* - t_i|) + \bar{d} + \theta \left(\frac{1}{2} + \frac{1}{2n}\right) \right]
\]

82
\[ + p(\chi_{\frac{1}{2} \sigma^L + \frac{1}{3} \sigma^R} = 2e) \left[ u(|t_L - t_i|) - \theta \left( \frac{1}{2} + \frac{1}{2n} \right) - u(|t_{LE}^* - t_i|) + \bar{d} + \theta \left( \frac{1}{2} - \frac{1}{2n} \right) \right] \]

\[ = \sum_{j=1}^{e-1} p(\chi_{\frac{1}{2} \sigma^L + \frac{1}{3} \sigma^R} = j) \left[ + \frac{2\theta}{n} \right] + \sum_{j=e+1}^{2e-1} p(\chi_{-i} = j) \left[ - \frac{2\theta}{n} \right] \]

\[ + \left( \frac{1}{2} \right)^{2e} \left[ u(|t_L - t_i|) - u(|t_{LE}^* - t_i|) + u(|t_{RE}^* - t_i|) - u(|t_R - t_i|) \right] \]

\[ + p(\chi_{\frac{1}{2} \sigma^L + \frac{1}{3} \sigma^R} = e) \left[ u(|t_{LE}^* - t_i|) - u(|t_{RE}^* - t_i|) \right] \]

\[ = \left( \frac{1}{2} \right)^{2e} [K_2 + K_3] + p(\chi_{\frac{1}{2} \sigma^L + \frac{1}{3} \sigma^R} = e) K_1, \]

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 \(\chi_{\frac{1}{2} \sigma^L + \frac{1}{3} \sigma^R}\) 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.
Proof of Proposition 5

(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

\[ \phi^C_i(\sigma^R, \sigma^L, \sigma^-_i) - \phi^C_i(\sigma^L, \sigma^R, \sigma^-_i) > 0 \quad \forall i \in \Omega, \]

since in this case, \( \sigma^R_i \neq \sigma^L_i \) for every \( i \in \Omega \). We calculate (directly or using Lemma 1):

\[ \mathbb{E}[U_i(R, \sigma^R_i)] = u(|t_R - t_i|) - \theta \left( \frac{1}{2} + \frac{1}{2n} \right), \]

\[ \mathbb{E}[U_i(L, \sigma^-_i)] = u(|t^w_L - t_i|) - \theta \left( \frac{1}{2} - \frac{1}{2n} \right) - d, \]

\[ \mathbb{E}[U_i(L, \sigma^L_i)] = u(|t_L - t_i|) - \theta \left( \frac{1}{2} + \frac{1}{2n} \right), \]

\[ \mathbb{E}[U_i(R, \sigma^-_i)] = u(|t^w_R - t_i|) - \theta \left( \frac{1}{2} - s_E + \frac{1}{2n} \right) - d, \]

and

\[ \phi^C_i(\sigma^R, \sigma^L, \sigma^-_i) - \phi^C_i(\sigma^L, \sigma^R, \sigma^-_i) = u(|t_R - t_i|) - u(|t_L - t_i|) - \frac{\theta}{n} + s_E \theta. \]

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

\[ \phi^C_i(\sigma^L, \sigma^R, \frac{1}{2} \sigma^L + \frac{1}{2} \sigma^R) > 0 \quad \forall i \in \Omega. \]

For any \( i \in \Omega \), \( \frac{1}{2} \sigma^L + \frac{1}{2} \sigma^R = B^2_{i/2} \). Thus by Lemma 1,

\[ \phi^C_i(\sigma^L, \sigma^R, \frac{1}{2} \sigma^L + \frac{1}{2} \sigma^R) = \left( \frac{1}{2} \right)^{2e} \left[ u(|t_L - t_i|) - u(|t_R - t_i|) - s_E \theta - \frac{\theta}{n} \right] + \frac{\theta}{n} \]

We argue as in the proof of Proposition 4.(iv). If \( e \geq e(n) : = \left\lfloor \frac{2n}{\theta} \right\rfloor \), then the \( (\frac{1}{2})^{2e} \) is dominated by \( \frac{\theta}{n} \) for large enough \( n \), so that indeed \( \phi^C_i(\sigma^L, \sigma^R, \frac{1}{2} \sigma^L + \frac{1}{2} \sigma^R) > 0 \forall i \in \Omega \), since uniformly in \( i \):

\[ u(|t_L - t_i|) - u(|t_R - t_i|) - s_E \theta - \frac{\theta}{n} \geq u(|t_L - t^R_{i_R}|) - u(|t_R - t^R_{i_R}|) - s_E \theta - \frac{\theta}{n}. \]
We proceed to calculate $\phi_i^C (\sigma_i^L, \sigma_i^R, (1-\varepsilon)\sigma_i^L_{\bot} + \varepsilon\sigma_i^R_{\bot})$. We have $\chi_{-i}^{(1-\varepsilon)\sigma^L + \varepsilon\sigma^R} = B_{1-\varepsilon}$. Thus, by Lemma 1:

$$
\phi_i^C (\sigma_i^L, \sigma_i^R, (1-\varepsilon)\sigma_i^L_{\bot} + \varepsilon\sigma_i^R_{\bot})
= \varepsilon^{2\varepsilon} \left[ u(|t_{RE}^w - t_i|) - u(|t_R - t_i|) - \bar{d} \right] \\
+ (1 - \varepsilon)^{2\varepsilon} \left[ u(|t_L - t_i|) - u(|t_{RE}^w - t_i|) \right] + \bar{d} - \left( s_E + \frac{1}{n} \right) \theta + \frac{\theta}{n} \\
= \varepsilon^{2\varepsilon} K_1(t_i) + (1 - \varepsilon)^{2\varepsilon} K_2(t_i) + \frac{\theta}{n},
$$

where we let

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

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 Corollary (Corollary 1), one sees that

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

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 $f_i(\varepsilon) := \phi_i^C (\sigma_i^L, \sigma_i^R, (1-\varepsilon)\sigma_i^L_{\bot} + \varepsilon\sigma_i^R_{\bot})$ is monotonically decreasing in $\varepsilon$, 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:

$$
\frac{\partial}{\partial t_i} f_i(\varepsilon) = 2e \left\{ \varepsilon^{2e-1} K_1(t_i) - (1 - \varepsilon)^{2e-1} K_2(t_i) \right\} \\
\frac{\partial^2}{\partial t_i^2} f_i(\varepsilon) = 2e(2e-1) \left\{ \varepsilon^{2e-2} K_1(t_i) + (1 - \varepsilon)^{2e-2} K_2(t_i) \right\}.
$$
Analogously to the proof of Proposition 4.(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 \( K_2(t_i) < 0, \frac{\partial^2}{\partial t^2} f_i(\varepsilon) < 0 \) \( \forall \varepsilon \in [0, 1] \), both \( \frac{\partial}{\partial t} f_i(\varepsilon) \) and \( f_i(\varepsilon) \) may have only one root in \([0, 1]\). But together with

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

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

END OF DOCUMENT
References


