ISSN 2282-6483

Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

Elections and Political Polarisation: Challenges for Environmental Agreements

Sarah Spycher

Quaderni - Working Paper DSE N° 1196

Elections and Political Polarisation: Challenges for Environmental Agreements

Sarah Spycher

Department of Economics, University of Bologna
spychersarah@gmail.com

This version: November 2024

Abstract

This paper examines the role of domestic elections and political polarisation in shaping international environmental agreements and how electoral dynamics may explain the limited success of current climate cooperation. I focus on two key factors: the impact of domestic electoral pressure on international policy decisions and the mismatch between short election cycles and long-term treaty commitments. Using a 4-stage game modelling a bilateral environmental agreement, I analyse how incumbents strategically balance policy preferences with reelection prospects. Results show that while a *green* incumbent is often forced to temper their ambitions, a *brown* incumbent faces fewer electoral constraints, explaining why stringent policies are harder to achieve. Nonetheless, electoral pressure can moderate policies, producing outcomes more aligned with the preferences of the median voter. Finally, I discuss how political polarisation, particularly in two-party systems, adds complexity to international cooperation on global public goods.

Keywords: international climate policy, political economy, elections, political polarisation, environmental policy making, public goods, externalities

---

I would like to thank Jean-Michel Benkert, Philipp Brunner, Nadia Ceschi, Simon Dietz, Niko Jaakkola, Igor Letina, Ralph Winkler and participants at the AURÖ Nachwuchsworkshop 2021 (Oldenburg), SAEE Workshop 2021 (Zurich), SURED 2022 (Ascona), EAERE 2022 (Rimini), EPS Forum 2022 (Graz), Environmental and Energy Economics Workshop 2023 (Bern), SSES Annual Congress 2024 (Lucerne), and IIPF Annual Congress 2024 (Prague), as well as seminar participants at the University of Bern, the London School of Economics and the University of Bologna for valuable comments. This research was supported by a SNSF Doc.Mobility Fellowship (P1BEP1_199981) and a SNSF Postdoc.Mobility Fellowship (P500PS_214320).

Non-technical summary

This paper examines how elections and political polarisation impact international environmental agreements, offering a potential explanation for the modest success of global climate cooperation. The study focuses on two main features of international environmental policy: first, policymakers in transnational negotiations often face domestic electoral pressures, shaping decisions such as to increase their re-election chances. Second, environmental treaties typically extend beyond a single government's term, creating a timing mismatch, as treaties are long-term while election cycles are short. This also means that treaty negotiation and ratification may fall to different administrations, potentially with conflicting policy goals depending on the level of political polarisation.

I use a four-stage game to model two countries considering setting up an emissions reduction treaty, concentrating on the political dynamics in one country with electoral competition. The current government negotiates the treaty while weighing its impact on the upcoming domestic election. After the election, the newly elected government – potentially from a different party – decides on ratification, followed by emissions decisions based on that outcome. I provide a detailed analysis of each stage in the game, ultimately characterising the resulting equilibrium treaties. In the model, the two political parties in the first country have differing environmental preferences: a *green* party that prioritises climate goals and a *brown* party with less willingness to engage in climate action. Political polarisation is defined by the distance between each party's environmental preferences and those of the median voter. The paper explores different scenarios based on which party is initially in power.

The results suggest that an incumbent government may intentionally propose a treaty that differs from what it would suggest without election pressures, in order to improve its re-election chances. This is due to a trade-off faced between a party's own policy preferences and its desire to win elections, resulting in policies that appeal to the median voter. The study identifies two possible treaty types: *consensus* treaties, designed to secure ratification regardless of the succeeding party, and *differentiation* treaties, where only one party would ratify, offering voters a clear contrast in policy approaches.

The findings reveal that brown incumbents can achieve consensus across a wide range of polarisation levels. Green incumbents, however, may need to moderate their ambitions to remain electable, especially in highly polarised settings. This asymmetry hints at a reason for limited climate cooperation: green parties face electoral pressure to temper ambitious plans, while brown parties are not similarly pressured to adopt stricter policies. Nevertheless, electoral pressure often nudges policies towards the middle, aligning them more closely with the median voter's preferences than policies created without election concerns.

1 Introduction

Anthropogenic climate change is widely recognised as one of the major global environmental issues of our times. In the past few decades, the international community has addressed the subject by negotiating many international environmental agreements (IEAs), most recently resulting in the *Paris Agreement* in 2015. However, little progress on climate change mitigation can be observed: the current pledges as agreed upon in the Paris Agreement are not ambitious enough to meet the recognised policy goal of keeping the increase in average surface temperature well below 2°C compared to pre-industrialised levels. In addition, in almost all countries, current greenhouse gas (GHG) emissions are above the pledged path (UNEP 2023).

Transnational cooperation on climate change mitigation poses a fundamental challenge to the international community: both the Paris Agreement and its predecessor the *Kyoto Protocol* indisputably demonstrate the difficulties of achieving ambitious environmental agreements as well as the reluctance of participating countries to comply with emission targets agreed upon. This lack of success is not surprising from an economic point of view: on the one hand, mitigation of anthropogenic climate change is impeded by the public goods property of GHG emission reductions. Each country's efforts to reduce emissions benefits all countries in a non-exclusive and non-rival manner, while costs are borne domestically. At the same time, no supranational authority exists that might enforce an efficient outcome. We therefore observe a global underprovision of emission reductions.

In this paper I investigate the role that domestic elections and political polarisation play for IEAs and to what extent they might be an explanatory factor for the modest success of current international cooperation on climate change mitigation. I focus on two key characteristics of international policy: first, agents involved in international negotiations are often subject to domestic electoral concerns and therefore, policy decisions might affect their chances of reelection in upcoming elections. Second, international treaties usually last beyond a government's incumbency. This, on the one hand, leads to a temporal disparity in the sense that environmental treaties are generally devised to last over a long period of time, while election cycles are comparably short. On the other hand, this implies that the negotiation and the ratification decision might be made by two different entities, which depending on the level of polarisation, might pursue very distinct environmental policy goals.

A good example of such deliberations is the behaviour of the US during the negotiations for the Kyoto Protocol. Al Gore, serving as vice president in the Clinton administration, participated in negotiating what was considered an ambitious target from the US perspective. However, the administration was fully aware that the Senate would likely reject the ratification of such a treaty. One could argue that Gore, anticipating his presidential run in the upcoming election, strategically positioned himself on environmental policy to bolster his electoral prospects. His campaign was unsuccessful, and the newly elected president, George W. Bush, chose not to ratify the Kyoto Protocol. This example underscores the rationale for separating negotiation and ratification decisions in the model, as they may be undertaken by different actors.

With these considerations in mind, I formulate a four stage game to model a bilateral environmental agreement, focussing on the strategic incentives arising from domestic electoral pressure. Two countries consider establishing a bilateral agreement on emission reductions, where the focus lies on political competition within country 1. In the first stage, the incumbent in country 1 negotiates a treaty, taking into account how its stringency might affect their chances of reelection in the upcoming domestic election. Following the election, which is stochastic and depends on the median voter's welfare, the new government decides whether to ratify the negotiated treaty. Finally, emission choices are made based on the ratification decision. I provide a detailed analysis of each stage in the game, ultimately characterising the resulting subgame-perfect equilibrium treaties.

Political competition within country 1 involves two rival parties: a *green* party and a *brown* party. The parties differ in their willingness to pay for environmental damage reduction, with one being more environmentally focused and the other less so, relative to the median voter. The preference distance of either party to the median voter is what is referred to as the level of political polarisation in the model. Throughout the paper, I will explore two distinct scenarios based on which party is in power at the start of the game. The environmental treaty is modelled as a cooperative agreement between the two countries to reduce emissions proportionally from the status quo. If the treaty is not ratified either by country 1 or 2, both countries default to the non-cooperative emissions level of the elected party.

I find that incumbent governments might indeed opt for a "suboptimal" treaty – relative to a scenario without election – to enhance their chances of reelection. This is influenced by various political economy factors and the degree of political polarisation. Incumbents face a fundamental trade-off between their policy preferences and their reelection prospects: choosing a treaty that aligns with their preferences might negatively impact their electoral chances, which can make it a profitable strategy to appeal to the median voter instead. This dynamic is reflected in the optimal treaty choices, which can be classified as either *consensus* or *differentiation* treaties. In a consensus treaty the incumbent anticipates the possibility of being replaced in the upcoming election and designs the agreement in a way that ensures its ratification by their potential successor, which is usually to the benefit of the median voter. In contrast, in a differentiation treaty, only one of the parties ratifies the treaty, leading to a situation where the two parties present differing environmental policies to voters in the election. This approach involves steering the treaty's ambition away from the challenging party's preferences, and consequently, from what the median voter would favour.

Equilibrium treaty outcomes reveal distinct pressures faced by green and brown incumbents in shaping climate policy. A brown incumbent can achieve consensus across a wide range of polarisation levels, staying true to their policy preferences with limited electoral concessions. By contrast, a green incumbent faces a much narrower set of conditions under which high-ambition policies can gain consensus. This asymmetry arises because the green party is generally willing to ratify most proposals from a brown incumbent, whereas a brown party will only agree to ratify a green incumbent's treaty under low levels of polarisation. These dynamics create a strategic bind for the green incumbent: while ambitious policies may reflect its platform, scaling back becomes essential to maintain electoral viability, especially when polarisation is high, and knowing that a loss could lead to a non-cooperative outcome under a brown successor – potentially more harmful than the status quo. These dynamics can help explain the persistence of modest climate cooperation success: the party pushing for ambitious policies is restrained by electoral dynamics, while the opposing party faces little pressure to engage in ambitious climate action. On the positive side, electoral pressure tends to moderate policies in a classic *convergence to the middle* manner, often resulting in welfare outcomes more closely aligned with the preferences of the median voter than would be achieved without elections.

Considering domestic political polarisation in the discussion of international cooperation is novel, and I demonstrate that this aspect crucially affects outcomes. This goes beyond the topic of environmental policy: I can illustrate how in a two-party system, the provision of a shared public good in general becomes more complex with political polarisation, an increasingly widespread phenomenon across diverse national contexts (Carothers and O'Donohue 2019). Furthermore, given the fact that the US, a prominent example of a two-party democracy, is one of the major global players when it comes to international cooperation, this connection will become increasingly relevant.

2 Related Literature

The question of how elections affect policy choices, and vice versa, has been widely discussed in contexts outside of international (environmental) cooperation. Persson and Tabellini

(1992) show that political processes such as elections may distort tax rate choices compared to what a social planner would do, while Besley and Coate (1998) highlight how fiscal policy investments can be used to influence future elections. Robinson and Torvik (2005) show how inefficient investments in local infrastructures might stem from attempts to influence elections. However, in Persson and Tabellini (1994), contrary to a majority of the public choice literature, the authors find that political incentives may also improve the equilibrium outcome through more credible commitment. Addressing how incumbent governments can influence policies of their successors, Alesina and Tabellini (1990) discuss the role of public debt as a means of limiting expenditures. My paper contributes to this literature by adding insights into the nexus of economic policy and political competition in the context of cross-border public goods provision, specifically in an environmental context.

The phenomenon of political polarisation has garnered significant attention across various disciplines due to its impact on policymaking in democratic systems. Esteban and Schneider (2008) argue that domestic polarisation diminishes countries' willingness to contribute to global public goods, such as international security. At the domestic level, polarised political systems and declining political diversity weaken governments' ability to provide public goods (Levin et al. 2021). Moreover, Baker et al. (2020) find that intense political polarisation in the US, particularly during closely contested elections, leads to spikes in economic policy uncertainty. Andreottola and Li (2024) analyse how mass polarisation, that is polarisation among the electorate, influences reform design. They argue that the incumbent has two viable strategies: either to increase their reelection chances or to design a reform that appeals to the opposition, thereby reducing the chance of repeal in case of election loss. This last result provides an interesting parallel to my finding of consensus and differentiation treaties, depending on the degree of polarisation and the size of the office rent. Few studies have examined the role of political polarisation in the context of environmental issues. Austen-Smith et al. (2019) show that polarised settings encourage inefficient environmental policies, as these are easier to reverse. Investigating national responses to changes in IEAs, Perrings et al. (2021) show that party polarisation weakens a country's commitment to these agreements. However, stakeholders who penalise parties for adopting extreme positions can moderate these effects, akin to the moderating influence of electoral pressure in my model.

In environmental economics, the theoretical analysis of self-enforcing IEAs often relies on non-cooperative game theory, particularly coalition formation games since the early 1990s (e.g., Carraro and Siniscalco 1993; Hoel 1992; Barrett 1994). These models typically involve two stages: countries first decide whether to join the agreement, and then signatories internalise emission externalities, while non-signatories act non-cooperatively (Wagner 2001). Such models generally yield pessimistic conclusions, predicting small coalitions and widespread free-riding, particularly when cooperation would yield large gains. This contradiction, known as the *paradox of international environmental agreements* (Kolstad and Toman 2005), arises because numerous large IEAs exist in practice. Finus and Maus (2008) address this by introducing the concept of *modest* IEAs, where only a fraction of emission externalities is internalised. They show that less ambitious agreements lead to larger coalition sizes, with the benefits of broader membership outweighing the higher emissions by individual members. Similar explanations for larger but less ambitious coalitions include Barrett (2002), Aldy et al. (2003), and Harstad (2022).

Most environmental economics models, including wide-ranging extensions, treat countries as homogeneous entities, represented by a single benevolent decision-maker, overlooking the potential interplay between domestic and international environmental policy (Finus 2008). To address this, a growing body of literature introduces hierarchical structures into models of international cooperation, incorporating insights from the political science literature, in which the relationship between domestic and international policy is described as a two-level game (Putnam 1988). This approach distinguishes between different governmental bodies within countries and emphasises the need to account for political economy factors like interest groups (Marchiori et al. 2017; Hagen et al. 2021), electoral concerns (Buchholz et al. 2005; Siqueira 2003), and domestic political structures (Loeper 2017). For instance, Spycher and Winkler (2022) show that by accounting for the hierarchical structure of international climate policy via the introduction of a strategic delegation stage, "broad-and-deep" agreements can be stabilised.

So far, only few papers combine international environmental cooperation with national political competition. Köke and Lange (2017) analyse the impact of ratification constraints in a strategic voting model, where they distinguish between "representatives" negotiating agreements and "pivotal agents" deciding on ratification. They formulate a coalition formation game and show how political dynamics within countries influence the size and scope of climate agreements, offering a public choice motivation to the findings of Finus and Maus (2008) by emphasising the role of political economy aspects in shaping international negotiations. Battaglini and Harstad (2020) model electoral considerations in climate negotiations, focussing on treaty design rather than participation. In a simple two-country framework, they show that incumbent governments may sign "weak treaties", that is treaties with too low a level of sanctions to guarantee compliance, in order to influence future elections in their favour. For instance, a government with low environmental preferences signs a treaty involving sanctions small enough for the median voter to favour non-compliance and thus reelection of the incumbent. This incentive is particularly strong when the benefits of staying in office are significant.

Coming more from a political science angle, Buisseret and Bernhardt (2018) explore how the prospect of electoral replacement affects international agreement terms. In their two-country model, renegotiation by the challenger is allowed post-election, and they find that ratification depends on the incumbent's stance towards the agreement and the timing of the election. If the incumbent is hostile (low environmental valuation), an agreement is signed only if the election is distant; if friendly, ratification happens only if the election is sufficiently near. Similarly, Melnick and Smith (2022) analyse the role of elections in bilateral agreement negotiations. In a model similar to mine but without externalities or political polarisation, they show how elections influence the types of treaties leaders are willing to sign, focussing on the resulting bargaining positions relevant for potential renegotiations. They find that *hawkish* incumbents, seeking to differentiate from challengers, may reject deals they would myopically accept, while for *dovish* incumbents maximising electoral prospects paradoxically cut better deals.

This paper builds on and complements model aspects of both Köke and Lange (2017) and Battaglini and Harstad (2020) while it aims at answering a question closely related to Buisseret and Bernhardt (2018) and Melnick and Smith (2022). My model enriches the bilateral treaty setup by Battaglini and Harstad (2020) along two main dimensions: First, in my setup the environmental externality goes both ways. Second, the setup is generally more nuanced, in that emission choices are modelled specifically as a result of reduction pledges, and that it allows for non-ratification (as opposed to compliance vs. non-compliance). In combination, this leads to the fact that on top of being able to replicate the results from Battaglini and Harstad (2020), additional and different equilibrium outcomes can be observed. Constrasting Melnick and Smith (2022), whose model endows the foreign country with agenda-setting power and considers a treaty merely about cost-sharing without a public goods characteristic, my model incorporates continuous policy choice, while abstracting from renegotiation. On top of that, all of the aforementioned papers largely abstract from political polarisation and its impacts on the degree to which international treaties are influenced.

Some aspects of the question at hand have also been investigated empirically. Cazals and Sauquet (2015) find that costly IEA ratifications are often delayed until after elections, although developing countries, where ratification costs are lower, may ratify before elections to gain indirect benefits, such as foreign aid. List and Sturm (2006) analyse the environmental policies of US governors and show that policy distortions depend on whether they are up for reelection and on the strength of the environmental lobby in their state. In *green* states, governors adopt less environmentally friendly policies when not seeking reelection, while in *brown* states, they become more pro-environment. They also find that lower political competition reduces policy manipulation. Both insights align closely with the findings of my model: the strategic incentive to seek consensus is reinforced by electoral pressures, while lower levels of polarisation enable a more genuine implementation of party agendas.

3 The Model

I consider two countries, country 1 and 2, which negotiate a bilateral environmental agreement on the levels of GHG emissions. In the status quo, both countries non-cooperatively choose emission levels such as to maximise the domestic welfare levels from the point of view of the respective incumbent. By doing so, due to the fact that GHG emissions are a transnational pollutant, countries do not take into account the negative externality their emission choice causes to the other country. The goal of the treaty is to commit to emission reductions relative to the status quo, and it depends on (i) who sits at the negotiation table and (ii) the political agenda put forward by the parties. We will analyse a situation in which a domestic election takes place in country 1 prior to the ratification of the treaty. It therefore is possible that the incumbent party, who negotiates the treaty, will be replaced in the election and thus, the challenging party will decide on the ratification of the agreement. Note, however, the challenger cannot renegotiate the treaty after winning the election. In the case of non-ratification, emission levels are chosen non-cooperatively by the elected government.

In the absence of domestic political competition, both countries are assumed to have identical environmental preferences. This allows us to isolate the impact of domestic political competition from any potential effects arising from differences between the two countries. Regarding negotiation dynamics, country 2 is modelled as a passive actor, playing no active role in shaping the treaty. Instead, it can only accept or reject the treaty proposed by country 1. More specifically, this reflects a *take-it-or-leave-it* offer from country 1, limited by country 2's participation constraint. See the extension in Section 5.3 for a relaxation of this assumption.

Note that in a political economy context, the socially optimal allocation depends on the distribution of environmental preferences across the varying actors involved.[1] Therefore, without imposing distributional assumptions, social welfare implications cannot be discussed. As a benchmark, I will thus consider the optimal treaty choice for the median voter, as well as the treaty choice of the party in power if they faced no election pressure.

3.1 Agency structure

In each country, domestic emissions e_i lead to benefits from productive activities according to a concave quadratic benefit function $B(e_i)$, while global emissions cause linear environmental

[1] Assuming a utilitarian social welfare function.

damages $D(E)$ with global emissions $E = e_1 + e_2$:

$$B(e_i) = \alpha e_i \left(\epsilon - \frac{1}{2} e_i\right), \qquad D(E) = \beta E, \tag{1}$$

where ϵ denotes the business-as-usual emissions, capturing the emission level if the economy ran at full capacity and no emission reductions were beneficial. The parameter α measures carbon efficiency, that is, how much GDP a country can produce per unit of domestic emissions and β indicates the level of environmental damage in monetary terms caused per unit of global emissions.

The environmental preference parameter θ captures an agent's valuation of environmental damage costs. We assume that in the absence of political economy considerations, the two countries' median voters share the same value of θ. We will focus on political competition within country 1 and thus assume that country 2 is represented by a government with median voter preferences. We therefore normalise $\theta_2 = \theta_{2,M} = \theta_{1,M} = 1$, where subscript M stands for median voter.

Within country 1, there are two competing parties: the party in power at the start of the game is the incumbent i, the other is the challenger j in the upcoming election. We assume that one party is "greener" (G) than the median voter and the other is "browner" (B), that is, they either have a higher or lower willingness to pay for environmental damage reduction than the median voter. Consequently, it holds that $\theta_{1,B} \leq \theta_{1,M} \leq \theta_{1,G}$. More precisely, the preference distance to the median voter is captured by the *polarisation* parameter ϕ:

$$\theta_{1,M} = 1, \qquad \theta_{1,G} = 1 + \phi, \qquad \theta_{1,B} = 1 - \phi. \tag{2}$$

This specification implicitly assumes that the the two parties are equally distanced from the median voter in terms of their environmental preference. An extension in Section 5.2 discusses asymmetric preferences and establishes the connection to the symmetric case.

Domestic welfare for any given agent $k \in \{M, B, G\}$ is defined by the difference between benefit and damage function, whereas the damage function is weighted with the agent's respective preference parameter θ:

$$W_{1,k}(e_1, e) = B(e_1) - \theta_{1,k} D(E), \tag{3a}$$

$$W_2(e_2, e) = B(e_2) - D(E). \tag{3b}$$

It follows that for a given value of global emissions, the welfare level of a greener agent (that is, with a higher value of θ) is always lower than that of a browner agent.

3.2 Treaty on Emission Reductions

The environmental treaty is modelled as the two countries cooperating with respect to emissions reductions. We consider an agreement design in which status quo emissions are reduced proportionally. The non-cooperative Nash equilibrium resulting from incumbents in both countries maximise domestic welfare as given by (3a) and (3b), taking the emission choice of the other country as given. Note that country 1's status quo depends on the incumbent's preference parameter $\theta_{1,i}$ (henceforth I will use the simplified notation θ_i):

$$e_{1,i}^{\text{sq}} = \epsilon - \frac{\beta}{\alpha}\theta_i, \qquad e_2^{\text{sq}} = \epsilon - \frac{\beta}{\alpha}. \tag{4}$$

When negotiating a treaty, the incumbent government $i \in \{B, G\}$ of country 1 suggests a treaty parameter $\delta_i \in [0, 1]$, specifying the amount of emission reduction in the agreement. There exists a *preferred* value for the treaty parameter in the absence of an election from the perspective of country 1's incumbent. However, i might want to suggest a different value, taking into account the upcoming election. Due to the fact that country 2 does not have any negotiation power, they are assumed to ratify any treaty that makes them at least indifferent to the non-cooperative outcome, that is, the outcome in the absence of an agreement. Note that i suggests a single parameter, meaning that the two countries both reduce their emissions by equal proportions. More formally, the incumbent i suggests a value δ_i, which determines agreement emission levels as follows:

$$\tilde{e}_{1,i} = \delta_i e_{1,i}^{\text{sq}}, \qquad \tilde{e}_2 = \delta_i e_2^{\text{sq}}.$$

The optimal treaty parameter in the absence of an election $\hat{\delta}_i$, henceforth called the *no-election treaty parameter*, is given as the solution to:

$$\begin{aligned} &\max_{\delta_i \in [0,1]} W_i(\delta_i, \theta_i) = B\left(\tilde{e}_{1,i}(\delta_i)\right) - \theta_i D\left(\tilde{e}_{1,i}(\delta_i), \tilde{e}_2(\delta_i)\right) \\ &\qquad \Rightarrow \quad \hat{\delta}_i(\theta_i) = \frac{1 + \beta\theta_i(\beta + \beta\theta_i - 3)}{(1 - \beta\theta_i)^2}. \end{aligned} \tag{5}$$

Note that the median voter, having other environmental preferences than the incumbent, has a different optimal value for δ_i, which is given as follows:

$$\begin{aligned} &\max_{\delta \in [0,1]} W_M(\delta, \theta_i) = B(\tilde{e}_{1,i}(\delta)) - \theta_M D(\tilde{e}_{1,i}(\delta), \tilde{e}_2(\delta)) \\ &\qquad \Rightarrow \quad \delta_M^*(\theta_i) = \frac{1 + \beta(\beta(1 + \theta_i) - 2 - \theta_i)}{(1 - \beta\theta_i)^2}. \end{aligned} \tag{6}$$

If the incumbent is green $\hat{\delta}_i < \delta_M^*$, and vice versa for a brown incumbent. Note that (6) does not indicate the median voter's optimal treaty choice in general, but the optimal treaty under a specific incumbent party, since it relates to i's status quo emissions (and not their own).

3.3 Timing

The game is formulated in four stages, with the timing given as follows:

1. **Agreement Stage**
   The two countries negotiate a bilateral agreement as described in Section 3.2. The treaty is then characterised by the parameter δ_i which maps into corresponding emission levels $\tilde{e}_1, \tilde{e}_2$ as defined by (8a) and (8b).
2. **Election Stage**
   An election takes place in country 1, where the median voter compares their welfare under the two parties. The reelection probability for the incumbent also stochastically depends on a relative popularity shock, as stated in Section 4.3.
3. **Ratification Stage**
   The election winner decides whether to ratify the agreement negotiated in Stage 1, which gives rise to ratification intervals of δ for both the incumbent and the challenger as detailed in Section 4.2.
4. **Emission Choice Stage**
   Domestic emission levels are chosen as a consequence of the ratification decision in Stage 3. In case of ratification, countries choose levels $\tilde{e}_1, \tilde{e}_2$, otherwise they set the non-cooperative emission levels $\hat{e}_1, \hat{e}_2$ as defined by (7a) and (7b).

The game can be illustrated by the following game tree:

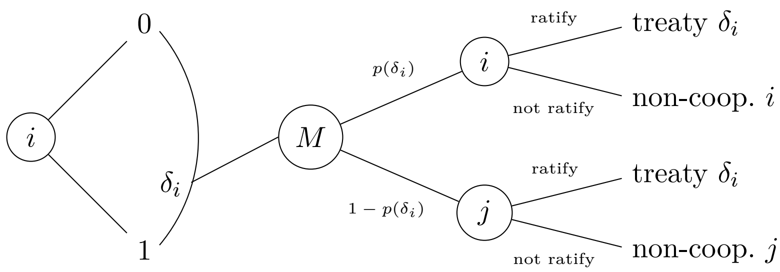


Figure 1: Decision tree within country 1

4 Solving the Model

The game is solved by backwards induction and we are looking for subgame perfect Nash equilibria. Hence, in this section equilibrium outcomes of each stage will be discussed depending on which party serves as the incumbent and on the degree of political polarisation.

4.1 Emission Choice Stage

In the emission choice stage, the government elected in Stage 2 chooses their country's emission level depending on whether they opted to ratify the agreement in Stage 3.

The alternative to ratification is the non-cooperative outcome. This means that country 2 and the elected government (which is either the incumbent i or the challenger j) in country 1 maximise domestic welfare (3a) and (3b) resulting in the following non-cooperative emission levels:

$$\hat{e}_1(\theta_h) = \epsilon - \frac{\beta}{\alpha}\theta_h, \qquad h = i, j \ , \tag{7a}$$

$$\hat{e}_2 = \epsilon - \frac{\beta}{\alpha}. \tag{7b}$$

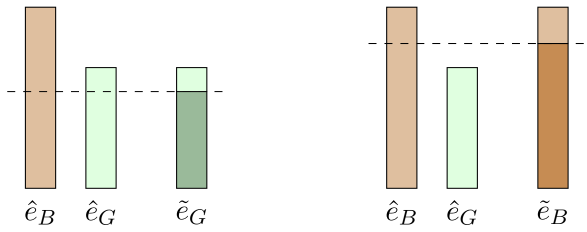
In case the treaty is ratified in Stage 3, the election winner sets emission levels such as to comply with the agreement negotiated in Stage 1, that is, to reduce status quo emissions according to δ_i. Emission levels are then given by:

$$\tilde{e}_1(\delta_i, \theta_i) = \delta_i \left(\epsilon - \frac{\beta}{\alpha}\theta_i \right), \tag{8a}$$

$$\tilde{e}_2(\delta_i) = \delta_i \left(\epsilon - \frac{\beta}{\alpha} \right). \tag{8b}$$

Importantly, note that if the challenger j wins the election and chooses to ratify the agreement, emission reductions will relate to the status quo emissions $e^{\text{sq}}_{1,i}$, that is, the level of $\tilde{e}_1$ is assumed to be fixed after Stage 1. Figure 2 sketches an exemplary scenario for both a green and brown incumbent.

This assumption seems intuitive thinking about a brown challenger who wins the election: if they choose to ratify, they adhere to the terms of the agreement as negotiated by the green incumbent, which results in lower emissions than what they would choose in the absence of a treaty. However, the assumption can seem problematic in the case of a green challenger who wins the election: if parties are highly polarised, it is possible that the green party's non-cooperative outcome is more ambitious than the negotiated treaty, as



(a) Green incumbent **(b)** Brown incumbent

Figure 2: Emission Level Scenarios

illustrated in Figure 2b. In this case, the green challenger would, under a ratified treaty, reduce emissions by less than what they would do non-cooperatively. A possible argument in favour of this assumption could be that the treaty negotiation essentially captures the setting up of an international permit market, where the negotiating party decides on the permit supply, which for some time horizon after the election is fixed. Another argument could be that if the green challenger were to voluntarily undercut treaty commitments, they would potentially forego the opportunity to negotiate a new and more ambitious treaty in the near future. Notwithstanding the underlying causes, this line of argument falls outside the scope of this model framework. However, in Section 5.1 I explore an extension in which the assumption of binding treaty emission is relaxed, allowing the treaty to serve merely as an upper bound, which the ruling party can undercut at their discretion.

Without loss of generality, we can normalise the problem by setting $\alpha = \epsilon = 1$. Also, throughout the paper, the range $\beta \in [0, 0.15]$ for the marginal damage parameter will be assumed. This is a very non-restrictive assumption, since $\beta = 0.15$, from a median voter perspective, would imply the worst case scenario of unmitigated climate change corresponding to approximately a 45% decrease in global GDP. While this is certainly much higher than what is commonly found to be realistic with respect to climate change in a global context (see, e.g., Hänsel et al. 2020), allowing for such high values might make sense from a more local perspective, for example for countries in Southeast Asia, where local impacts are expected to be significantly higher than the global average (see, e.g., Swiss Re 2021). In any case, allowing for such high values of β ensures that results are not driven by a too optimistic evaluation of environmental damages.

4.2 Ratification Stage

Whoever is in charge at the ratification stage, that is, the election winner in Stage 2, will decide whether to ratify the treaty on the table. As previously stated, we assume that renegotiation of the treaty is not available to the governing party after the election. This is a reasonable assumption in the context of this model, since treaty negotiations usually take place over a long time horizon and that it is not possible for a newly elected government of one country to immediately renegotiate an international agreement, as seen, for example, in the case of the US and the Kyoto Protocol. This feature of the model contrasts the setup of Buisseret and Bernhardt (2018) as well as Melnick and Smith (2022), where agreements made before an election serve as a starting point for any subsequent renegotiation.

4.2.1 Ratification Intervals

The incumbent i or the challenger j, when elected, ratify the agreement whenever:

$$\Delta W_h = W_h(\tilde{e}_1(\delta_i, \theta_i), \tilde{e}) - W_h(\hat{e}_1(\theta_h), \hat{e}) \geq 0, \qquad h = i, j \ . \tag{9}$$

This leads to two intervals for which elected party will ratify the agreement, each defined by the two threshold values for δ:

$$[\underline{\delta}^i, \bar{\delta}^i] = \left[\max\left\{0, \frac{1 + \beta\theta_i\,(\beta(2+\theta_i) - 4)}{(\beta\theta_i - 1)^2}\right\}, 1\right], \tag{10}$$

$$\begin{aligned}[\underline{\delta}^j, \bar{\delta}^j] = \Bigg[&\max\left\{0, \frac{1 - \beta\left[\theta_i - \theta_j(\beta + \beta\theta_i - 2)\right] - \sqrt{M}}{(\beta\theta_i - 1)^2}\right\}, \\ &\min\left\{\frac{1 - \beta\left[\theta_i - \theta_j(\beta + \beta\theta_i - 2)\right] + \sqrt{M}}{(\beta\theta_i - 1)^2}, 1\right\}\Bigg],\end{aligned} \tag{11a}$$

$$\text{where } M = \beta^2\theta_j(\beta - 1)\left[\theta_j(\beta + 2\beta\theta_i - 3) - 2\theta_i(\beta\theta_i - 1)\right] \tag{11b}$$

Note that (11b) has to be non-negative for (11a) to be well defined. Also, the incumbent's upper threshold value corresponds to no emission reductions, that is, to the non-cooperative outcome.

We will see in the following that it depends on which party is the incumbent and which is the challenger in order to state how these thresholds relate to each other. We define:

$$\Delta\underline{\delta} \equiv \underline{\delta}^i - \underline{\delta}^j, \tag{12a}$$

$$\Delta\bar{\delta} \equiv \bar{\delta}^i - \bar{\delta}^j, \tag{12b}$$

which define the respective order of the two parties' lower and upper ratification thresholds.

While the incumbent of country 1 suggests a treaty parameter δ_i, country 2 is assumed to not have any negotiation power. Still, country 1's treaty suggestion is limited by country 2's participation constraint, that is, country 2 must not be worse off than under no agreement, in which case they expect the incumbent's non-cooperative emission choice. The corresponding ratification thresholds for country 2 are then given as follows:

$$\Delta W_2 = W_2(\tilde{e}_2(\delta_i, \theta_i, \theta_2), \tilde{e}) - W_2(\hat{e}_2(\theta_2), \hat{e}) \geq 0 \tag{13}$$

$$\Leftrightarrow \quad [\underline{\delta}_2(\theta_i), \bar{\delta}_2(\theta_i)] = \left[\frac{1 + \beta(2\beta\theta_i + \beta - 4)}{(\beta - 1)^2},\ 1\right]. \tag{14}$$

Country 2's ratification interval only depends on the incumbents' preference parameter. This is due to the fact that their treaty partner in the agreement stage is the incumbent and even if the challenger were to win the election and ratify the treaty, country 2's emission reduction commitment only relates to the treaty signed with the incumbent. Implictly, we assume country 2 to not be sophisticated enough to anticipate the possibility of facing the challenger's non-cooperative outcome if they were elected and did not ratify.

The threshold values (10) and (11a) can be ordered resulting in a partition of ranges for the treaty parameter. The ordering will depend on which party is the incumbent and the size of the polarisation parameter ϕ. Four different cases can emerge:

(A) $\delta \in [\underline{\delta}^i, \bar{\delta}^i], \delta \notin [\underline{\delta}^j, \bar{\delta}^j]$: only the incumbent ratifies,

(B) $\delta \notin [\underline{\delta}^i, \bar{\delta}^i], \delta \in [\underline{\delta}^j, \bar{\delta}^j]$: only the challenger ratifies,

(C) $\delta \in [\underline{\delta}^i, \bar{\delta}^i], \delta \in [\underline{\delta}^j, \bar{\delta}^j]$: both ratify,

(D) $\delta \notin [\underline{\delta}^i, \bar{\delta}^i], \delta \notin [\underline{\delta}^j, \bar{\delta}^j]$: none ratify.

Note that technically, there is an additional case in which a treaty that would be ratified by at least one of the parties in country 1 but not by country 2. However, from a theoretical point of view there is no difference to case D regarding the resulting emission choices. Thus, henceforth, this scenario will be captured by case D.

In the following, we will discuss specific ratification intervals in the context of a green and a brown incumbent.

Green incumbent

In a first step, let us assume the green party serves as the incumbent. We can therefore assign preference parameters $\theta_i = 1 + \phi$ and $\theta_j = 1 - \phi$. Note that if the parties are too polarised, the challenger would never ratify a treaty suggested by the incumbent. This is the case if no real value δ_i satisfies (9), that is, whenever (11b) is negative. This holds true for polarisation values $\phi \leq \bar{\phi}^G(\beta)$, where $\frac{d\bar{\phi}^G}{d\beta} > 0$, as shown in Lemma A.1[2]. This implies that for higher environmental damages, a ratification interval of the brown challenger exists for higher degrees of polarisation.

We can now specify ratification intervals (10) and (11a) in case of a green incumbent as detailed in Proposition A.1, with corresponding comparative statics in Proposition A.2. We find that while the incumbent's upper threshold is independent of ϕ, their lower ratification threshold decreases with the distance from the median voter. Intuitively this means that a greener incumbent will be more willing to ratify strict treaties. For the brown challenger, the upper threshold decreases and the lower threshold increases in ϕ, meaning that their ratification interval becomes more narrow with increasing polarisation. Consequently, higher polarisation leads to a more narrow range of δ_i that would allow for ratification by both parties.

Proposition 1 (Ordering of Ratification Thresholds with $i = G$)

1. *In the case of the green incumbent and whenever $\phi \leq \bar{\phi}^G$, the ratification thresholds* (10) *and* (11a) *relate to each other as given in the following:*

$$\Delta\underline{\delta}^{i=G} \leq 0, \tag{15a}$$

$$\Delta\bar{\delta}^{i=G} \geq 0. \tag{15b}$$

2. *The lower ratification threshold of parties in country 1 relate to that of country 2 as follows:*

$$\underline{\delta}^{i=G} - \underline{\delta}_2 \leq 0, \tag{16a}$$

$$\underline{\delta}^{j=B} - \underline{\delta}_2 \geq 0. \tag{16b}$$

The first part of Proposition 1 states that the green incumbent will always sign stricter treaties than the brown challenger[3]. While no treaty is unambitious enough for the incumbent (technically they can negotiate a treaty with $\delta_i = 1$, which corresponds to their non-cooperative outcome), the challenger will not always sign such treaties. The reason for this is that if emission reductions are negligible, the positive effect of damage reductions (also via less externalities from country 2) does not outweigh the negative effect of not being able to choose emissions freely according to their own optimal non-cooperative outcome.

[2] Henceforth, intermediate results are relegated to Appendix A.
[3] The proofs for all propositions presented in the main text are provided in Appendix B.

The second part then states that country 1 with a green incumbent is willing to ratify more strict treaties than allowed for by country 2's participation constraint. This implies that in some cases, country 2's lower ratification threshold is binding, rather than that of the green incumbent. The brown challenger, conversely, ratifies less ambitious treaties than country 2.

This leads us to a discussion of which cases arise along the spectrum of possible treaty parameters, following Proposition 1:

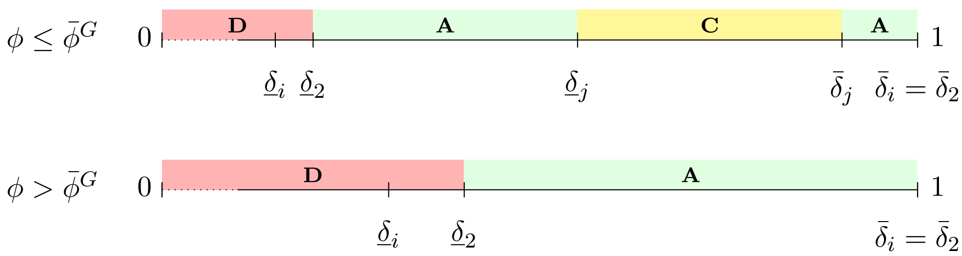


Figure 3: Green incumbent ratification thresholds

The two scenarios are distinguished by whether ratification thresholds for the challenger exist, following the threshold value $\bar{\phi}^G$ as defined by Lemma A.1. If the thresholds for the challenger do not exist, the feasible range for the treaty parameter is limited by country 2. Note that the colours in Figure 3 will henceforth indicate which parties ratify the treaty: green and brown for the two parties respectively, yellow for both parties and red for none.

Brown incumbent

Next, we consider the scenario in which the brown party is the incumbent and therefore environmental preference parameters are given by $\theta_i = 1 - \phi$ and $\theta_j = 1 + \phi$. Ratification intervals following (10) and (11a) are then stated in Proposition A.3.

We find that the lower ratification threshold for the brown incumbent increases in the distance from the median voter, whereas the upper threshold is independent of ϕ, as detailed in Proposition A.4. Intuitively, the browner the incumbent, the less strict the treaty can be for them to ratify. For the challenger, higher polarisation decreases both upper and lower thresholds. The greener the challenger, the more ambitious the treaty can be on the lower end and has to be on the upper end, for them to ratify. If a treaty is too weak, the damage reduction does not compensate for insufficiently low emission levels. This essentially means that their ratification interval shifts downwards.

Proposition 2 (Ordering of Ratification Thresholds with $i = B$)

1. *In the case of the brown incumbent, the ratification thresholds* (10) *and* (11a) *relate to each other as given in the following:*

$$\Delta\underline{\delta}^{i=B} \geq 0, \tag{17a}$$

$$\Delta\bar{\delta}^{i=B} \geq 0. \tag{17b}$$

2. *The incumbent's lower ratification threshold relates to that of country 2 as follows:*

$$\underline{\delta}^{i=B} - \underline{\delta}_2 \geq 0, \tag{18a}$$

$$\underline{\delta}^{j=G} - \underline{\delta}_2 \leq 0. \tag{18b}$$

The first part of Proposition 2 states that the green challenger will sign more ambitious treaties than the brown incumbent. However, at the upper end of the spectrum, there are very unambitious treaties that the incumbent will sign and the challenger not. Intuitively, these are contracts that are so unambitious in terms of damage reduction that the green challenger is better off with their non-cooperative outcome.

Secondly, Proposition 2 states that country 1 with a brown incumbent is willing to ratify less ambitious treaties than would be allowed for by country 2's participation constraint, and vice versa for the green challenger. Consequently, the following cases arise:

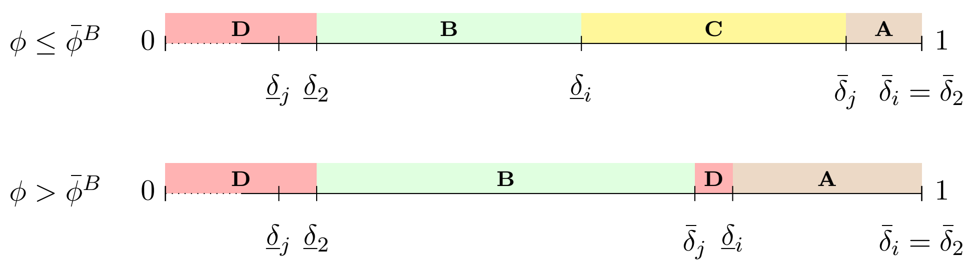


Figure 4: Brown incumbent ratification thresholds

The two scenarios are separated by whether ratification intervals of the incumbent and challenger overlap or not, that is, depending on $\underline{\delta}_i \lesseqgtr \bar{\delta}_j$. This condition yields a threshold value for polarisation $\bar{\phi}^B$ as defined in Lemma A.2. In case of no overlap, this means that no treaty parameter leads to ratification by both parties, as depicted in the second scenario (no area C) in Figure 4. Hence, if polarisation is very high, there exists an interval for the treaty parameter, which will not be signed by any of the two parties, since it is too ambitious for the brown incumbent while being not ambitious enough for the green challenger.

4.3 Election Stage

The domestic election is modelled as devised by Battaglini and Harstad (2020). The median voter faces the choice between the incumbent i and the challenger j and considers how each will affect their welfare: depending on the election outcome, country 1 can either (i) be part of the treaty and choose emissions as negotiated by i or (ii) live in a world without a treaty where both countries choose the non-cooperative outcome. In the former case, both parties will act identically whereas in the latter, the outside option differs between the two. The median voter can, as defined by the ratification intervals derived in Section 4.2, anticipate what the consequence of electing either of the two parties is.

The median voter's welfare difference between a government i and j is denoted by ΔW_M and the incumbent is consequently re-elected whenever:

$$\Delta W_M \equiv W_M^i - W_M^j \geq \Omega, \qquad \text{where } \Omega \sim U\left[-\frac{z}{\sigma}, \frac{1-z}{\sigma}\right]. \tag{19}$$

The parameter $z \geq 0.5$ quantifies an incumbency advantage, that is, the reelection probability for the incumbent in the absence of any policy differences between the two parties. The parameter σ captures the density of a popularity shock. A high value of σ (low variance) means that policy differences are more likely to dictate the outcome of the election, whereas low values of σ (high variance) increase noise and thus make random popularity shocks more important. The parameter can therefore also be interpreted as a value for policy salience. An example of such a shock could be an exogenous change in the political climate with respect to environmental issues, as for example, the Fukushima nuclear disaster in 2011. For reelection probabilities to be interior in $(0, 1)$, the variance in the popularity shock is limited to be $\sigma < \bar{\sigma}$ as defined in Lemma A.3, which however does not restrict the presented results in later sections. Proposition 3 gives a full characterisation of the reelection probability for the incumbent party in Stage 2, as a consequence of (19).

Proposition 3 (Stage 2: Reelection Probabilities)
Given that $\sigma < \bar{\sigma}$ and $z \geq 0.5$, reelection probabilities for cases A – D are defined by:

$$p^l(\delta_i) = \sigma \Delta W_M^l + z \qquad l \in \{A, B, C, D\}, \tag{20}$$

with

$$\begin{aligned}
\Delta W_M^A &= W_M(\tilde{e}_1(\theta_i, \delta_i), \tilde{e}) - W_M(\hat{e}_1(\theta_j), \hat{e}) && \textit{(only the incumbent party ratifies)} \\
\Delta W_M^B &= W_M(\hat{e}_1(\theta_i), \tilde{e}) - W_M(\tilde{e}_1(\theta_i, \delta_i), \hat{e}) && \textit{(only the challenging party ratifies)} \\
\Delta W_M^C &= 0 && \textit{(both parties ratify)} \\
\Delta W_M^D &= W_M(\hat{e}_1(\theta_i), \hat{e}) - W_M(\hat{e}_1(\theta_j), \hat{e}) && \textit{(none of the parties ratify).}
\end{aligned}$$

This reelection probability is a function of the treaty parameter δ_i that determines which case $A - D$ emerges. Therefore, the reelection probability between cases differs, since W_M^i and W_M^j depend on whether ratification occurs in Stage 3. Straightforwardly, the median voter's welfare level is affected in the cases where the incumbent and the challenger will take different ratification decision (A and B). In the case where both parties will ratify C, this is not the case because the challenger is tied to the treaty negotiated by the incumbent. In the last case D, again the median voter's welfare levels are different since the two parties will choose differing non-cooperative emission levels. Note that the reelection probability is a function of the treaty parameter in cases A and B but not in cases C and D.

4.4 Agreement Stage

The incumbent government negotiates an agreement such that their expected welfare is maximised:

$$\max_{\delta_i} \; p(\delta_i)\Big[W_i(\text{'i in power'}) + R\Big] + (1 - p(\delta_i))\Big[W_i(\text{'j in power'})\Big], \tag{21}$$

and $W_i(\cdot)$ and $p(\delta_i)$ depend on cases $A - D$, as detailed in the previous sections.

R denotes the rent from staying in office, which can capture any inherent benefits from staying in power. Battaglini and Harstad (2020) refer to R as an indirect measure for political polarisation: the further apart the two parties, the more important holding the office is, for example, to influence domestic policy unrelated to emission choice. Furthermore, the level of office rents might differ between political systems: presidential systems would then be associated with higher values of R, as opposed to parliamentary systems, in which the surplus from being in office is more spread out across political actors and being in office comes with less power to push one's own agenda.

The incumbent's objective function illustrates the fundamental trade-off that they face: choosing the treaty which maximises their welfare function when in power, that is $\hat{\delta}_i$, might not be optimal when considering the effect this choice has on the reelection probability. It can therefore be a profitable strategy to adjust the treaty parameter such as to influence reelection prospects as well as the welfare level in case of election loss in a favourable way. The incentive to increase the reelection probability is particularly strong when the office rent is high: in that case, the relative weight of the actual policy choice is reduced and staying in power becomes more important.

When solving this trade-off, the incumbent government chooses δ_i, perfectly anticipating which case will materialise according to the derived ratification intervals. Due to the fact that (21) is a non-smooth function, they compute expected maximum welfare levels for each case A - D and then opt for the case yielding the highest expected welfare and the corresponding optimal treaty in that range.

Note that in the range in which both parties would ratify the agreement, that is, case C, it is welfare-maximising for the incumbent to set $\delta_i^* = \hat{\delta}_i$. This is due to the fact that the objective function qualitatively corresponds to the maximisation problem in the absence of an election, that is, (5):

$$\begin{aligned} W_i^C &= p^C \Big[W_i(\tilde{e}_1(\theta_i, \delta_i)) + R \Big] + (1 - p^C) \Big[W_i(\tilde{e}_1(\theta_i, \delta_i)) \Big] \\ &= W_i(\tilde{e}_1(\theta_i, \delta_i)) + zR. \end{aligned} \tag{22}$$

Intuitively, since the reelection probability is not influenced by the choice of agreement, no distortion of policy choice is necessary, therefore allowing for the *first-best* outcome to materialise. However, we will see that the treaty parameter $\hat{\delta}_i$ does not lie in the range of case C when polarisation becomes too large.

Optimal treaty choices can be categorised into two groups: consensus treaties, which are ratified independent of the election outcome and differentiation treaties, which are only ratified by either the incumbent or the challenger. Within the two groups, treaty types differ in terms of the incumbent's underlying rationale, as detailed in the following:

- **Consensus treaty:** ratified by both (case C)
  - *First-best* (FB): optimal treaty is equivalent to no-election treaty
  - *Compromise* (COMP): treaty ambition is shifted towards challenging party's preferences to ensure ratification independent of election outcome
- **Differentiation treaty:** ratified by either incumbent (case A) or challenger (case B)
  - *Distinction* (DIST): treaty ambition is shifted away from challenging party's preferences to stress policy differences towards median voter

– *Assimilation* (ASSIM): treaty ambition is shifted towards median voter preferences to improve electoral prospects

– *Insurance* (INS): treaty ambition adapated such as to ensure an acceptable outcome in case of election loss

The respective availability of these treaty types depends on polarisation levels and will be defined in detail later in this section. For illustrative purposes, the optimal treaty choice for the incumbent is henceforth presented in numerical examples. The following parameters will be assumed throughout:

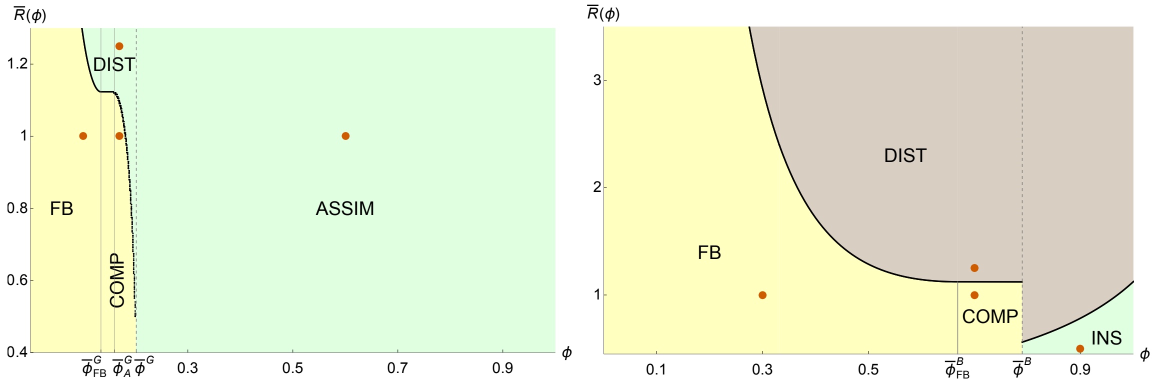
$$z = 0.55, \quad \beta = 0.05, \quad \sigma = 0.8.$$

All of these parameter values are unexceptional, in that they do not drive any of the results presented, and postulate (i) a 5 percentage point incumbency advantage, which following, for example, Gelman and King (1990) and Levitt and Wolfram (1997) are a middle-of-the-road estimate, (ii) environmental damages in the absence of any policies pursuing climate change mitigation would constitute approximately an 18% reduction of GDP, and (iii) a shock density parameter to mirror environmental policy being relatively salient such that 80% of policy differences between the contenders transmit into reelection probabilities, mirroring the currently high visibility of the climate crisis in policy debates in many countries.

In Figure 5 the shaded areas indicate which party ratifies the treaty: yellow for both, green and brown for the respective parties. The orange dots refer to the specific numerical scenarios which will be discussed in detail in Sections 4.4.1 and 4.4.2. Figure 5a illustrates optimal treaty choices by a green incumbent depending on the level of polarisation and the office rent.[4] A common ratification interval only exists for relatively low levels of polarisation. In this area, the treaty can be of the type *first-best* or *compromise* for sufficiently low levels of the office rent, depending on the availability. For a higher office rent, the "gamble" of a *distinction* treaty is worthwhile because of the increased importance of reelection. For higher levels of polarisation, no treaty will lead to ratification by the challenger and therefore the sole focus lies on appealing to the median voter by choosing an *assimilation* treaty.

Analogously, Figure 5b shows the optimal treaty choices for a brown incumbent. Common ratification is optimal for a large range of polarisation levels, as is the treaty type *first-best* for a sufficiently low office rent. The level of the office rent which separates *distinction* from *first-best* or *compromise* treaties decreases in polarisation, since consensus becomes more costly the more distinct party preferences are and therefore reelection becomes relatively

[4] Note that this is an illustrative example including a numerical approximation for $\bar{R}(\phi)$ in the range $\phi \in (\phi_A^G, \phi_{FB}^G)$. A more detailed discussion follows later in this section.



(a) Green incumbent treaty choices

(b) Brown incumbent treaty choices

Figure 5: Optimal treaty choices

more important. For high levels of polarisation and low levels of the office rent, *insurance* treaties are optimal: common ratification is no longer possible but the relative weight of the resulting environmental policy is high and thus the incumbent aims at guaranteeing an acceptable policy outcome even in case of election loss. This rationale becomes increasingly relevant as polarisation intensifies since the non-cooperative outcome grows more extreme, as demonstrated by the rise in $\bar{R}$ within this range.

In the remainder of this section, optimal treaty choices for both green and brown incumbents under varying degrees of polarisation will be discussed as well as formally specified. Furthermore, the results will be complemented by corresponding numerical illustrations, to allow for direct comparison between different treaty types in terms of emission levels and reelection probabilities. Also, a natural question arises regarding a ranking of the presented treaty types in terms of implications on welfare. Given the agency structure of the model, this is not an obvious exercise. The approach chosen here focusses on the median voter and presents two distinct measures to allow for treaty comparison.

First, the measure for *treaty fit* quantifies how the median voter judges the resulting treaty ambition in comparison to the incumbent's optimal no-election choice. The value ΔW_M^{TF} specifically captures the percentage difference in median voter welfare between an implemented optimal and no-election treaty:

$$\Delta W_M^{TF} = \frac{W_M(\hat{\hat{\delta}}_i(\theta_i)) - W_M(\delta_i^*(\theta_i))]}{W_M(\hat{\delta}_i(\theta_i))} \tag{23}$$

Second, the measure for *political fit* captures the median voter welfare consequence a treaty has in a given political environment. Specifically, it gives the percentage difference in expected median voter welfare between what would happen under the no-election and the optimal treaty:

$$\Delta W_M^{PF} = \frac{\mathbb{E}[W_M(\hat{\delta}_i(\theta_i))] - \mathbb{E}[W_M(\delta_i^*(\theta_i))]]}{\mathbb{E}[W_M(\hat{\delta}_i(\theta_i))]} \tag{24}$$

Essentially, these two metrics allow us to capture two key insights: the first shows how much the median voter, deterministically, likes a given optimal treaty depth compared to the incumbent's treaty choice without electoral constraints. The second measure then reflects how their welfare is affected in expectation: given that the incumbent internalises electoral pressure, they account for the possible ratification outcomes, which might differ between the optimal and the no-election treaty. For a first-best treaty, both measures are zero: the incumbent chooses the same treaty as in the absence of an election and hence the median voter faces the same (expected) welfare. This is not the case for all other treaty types.

Finally, the numerical illustration seeks to contextualise the size of the office rent. Essentially, R captures the relative importance of climate policy with respect to other policy aspects. Intuitively, if R is high, this means the party in office draws high levels of welfare from aspects unrelated to the climate policy choice. In order to capture this relative importance in a comparable way, the following ratio is introduced, where W_i^l corresponds to the welfare of the party in office for the case that follows from the optimal treaty parameter:

$$\omega(R) = \frac{W_i^l(R=0)}{W_i^l(R=0)+R} \qquad \text{for } l \in \{A, B, C, D\}. \tag{25}$$

As an extreme example, $R = 0$ can thus be interpreted as climate policy making up 100% of an administration's welfare. As R increases, the importance of climate policy and thus the value of ω decreases.

4.4.1 Green incumbent

In this section, I present optimal treaty outcomes when the incumbent party is green, depending on the degree of political polarisation. Ratification thresholds and thus the partition of the δ-range into the four cases are given as depicted in Figure 1, and corresponding re-election probabilities are anticipated as given in Section 4.3.

The degree of polarisation can be broadly classified into two categories – low polarisation and high polarisation – differentiated by the presence or absence of a common ratification interval and captured by the threshold value $\bar{\phi}^G$ as defined by (A.1). Whether the consensus treaty can be first-best or will be a compromise treaty depends on whether the no-election treaty parameter lies within area C, the threshold parameter for which is given by $\bar{\phi}^G_{FB}$. Details on the ranking and properties of these two threshold parameters are given in Lemma A.4. Figure 6 summarises polarisation ranges and thresholds for a green incumbent.

Figure 6: Green incumbent polarisation ranges

Proposition 4 details the possible treaty outcomes for a green incumbent. We first consider the case in which polarisation is low, that is when common ratification is possible. The green incumbent trades off the importance of environmental policy outcome as a consequence of the treaty parameter versus the importance of staying in office due to the office rent. The threshold value for the office rent, which separates whether the incumbent prioritises the former or the latter, is a function of polarisation and is given by:

$$\bar{R}^G(\phi) = \begin{cases} \arg\min |W^C(\hat{\delta}_i) - W^A(\underline{\delta}_j)| & \text{for } \phi \leq \phi^{\bar{G}}_{FB} \\ \arg\min |W^C(\underline{\delta}_j) - W^A(\underline{\delta}_j)| & \text{for } \bar{\phi}^G_{FB} < \phi \leq \bar{\phi}^G_A \\ \arg\min |W^C(\underline{\delta}_j) - W^A(\delta^*_{i,A})| & \text{for } \bar{\phi}^G_A < \phi \leq \bar{\phi}^G. \end{cases} \tag{26}$$

If the office rent lies above the threshold value as defined by (26), the incumbent will choose differentiation. The first line refers to the case when the first-best treaty is available, in which case the two potential treaty outcomes differ in ambition ($\hat{\delta}_i$ or $\underline{\delta}_j$). As polarisation increases, the resulting treaty parameters will only differ marginally, and purely serve the cause of differentiation. Again, in the small range $\phi \in (\bar{\phi}^G_{FB}, \bar{\phi}^G_A)$, treaty ambition differs because the global maximum of the W^A function, i.e. $\delta^*_{i,A}$, is available for the differentiation treaty, while the compromise treaty remains unchanged at $\underline{\delta}_j$. Note that in both, that is, in case of a compromise and a distinction treaty, the optimal choice by the green incumbent is an agreement that is weaker than what they would prefer in the absence of an election.

For higher levels of polarisation, ratification by both parties never occurs and a departure from the incumbent's no-election treaty is purely motivated by a desire for reelection, therefore moving closer to the median voter's preferred treaty.

Proposition 4 (Treaty Outcomes for $i = G$)

(i) For low levels of polarisation, that is for $\phi \leq \bar{\phi}^G$, it holds that:

1. *it is never optimal for the green incumbent to choose a treaty in the upper area A, that is, $\delta \in [\bar{\delta}_j, 1]$,*

2. the green incumbent chooses a consensus treaty if $R \leq \bar{R}^G(\phi)$ and a distinction treaty (lower area A) otherwise.

3. The chosen consensus treaty is first-best for $\phi \leq \bar{\phi}^G_{FB}$, that is $\delta^*_i = \hat{\delta}_i$, otherwise it is a compromise treaty with $\delta^*_i = \underline{\delta}_j + \epsilon$.

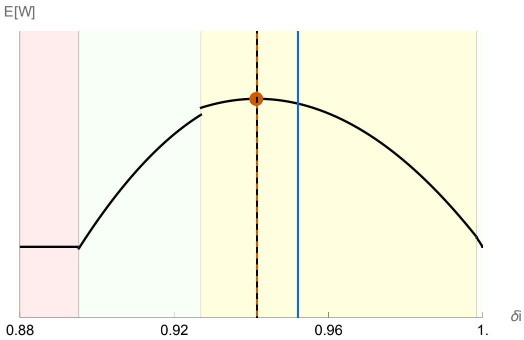
4. For $\phi \in (\bar{\phi}^G_{FB}, \bar{\phi}^G_A)$, the distinction treaty only marginally differs from the compromise treaty, with $\delta^*_i = \underline{\delta}_j - \epsilon$. For $\phi \in (\bar{\phi}^G_A, \bar{\phi}^G)$, the incumbent chooses $\delta^*_{i,A}$, that is, the global maximum within the lower area A.

(ii) For higher levels of polarisation, that is for $\phi > \bar{\phi}^G$, it is optimal for the green incumbent to choose a treaty with $\delta^*_i \in (\hat{\delta}_i, \delta^*_M)$, which is an assimilation treaty.

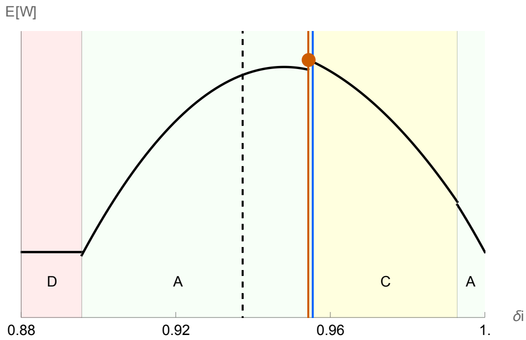
In the following, the incumbent's expected welfare level W_i will be plotted against the agreement parameter δ_i. The shaded backgrounds indicate the case $A - D$ the respective value of δ_i would give rise to. The dotted line shows the value of $\hat{\delta}_i$ as given by (5), that is, the preferred treaty parameter in the absence of an election, and the orange line indicates the optimal parameter choice as a solution to (21). The blue line depicts the preferred treaty parameter for the median voter (relating to the incumbent's non-cooperative emissions) as given by (6). Figure 7 illustrates the outcomes of treaties across various exemplary political environments, each defined by specific levels of polarisation and office rent.

In Figure 7a an example of a first-best treaty is depicted, in which the no-election treaty emerges. Figure 7b illustrates a case in which the office rent is below $\bar{R}^G$ and the level of polarisation does not allow for a common ratification of the no-election treaty. Reducing the ambition of the treaty with respect to the no-election treaty is profitable, as the green incumbent prefers a weaker but ratified treaty for sure over the risk of ending up with the challenger's non-cooperative outcome. Figure 7c contrastingly illustrates the case in which the green incumbent is willing to take a risk: reelection chances are increased by forcing a differentiation against the challenger, in which case the median voter prefers an agreement that is deemed too strict over the non-cooperative outcome by the brown challenger, that is $\Delta W^A_M > 0$ and hence $p^A(\delta^*_i) > p^C$. This increased reelection probability is, due to the prospect of an office rent above $\bar{R}^G$, worth more than the loss from an undesirable outcome in case of election loss.

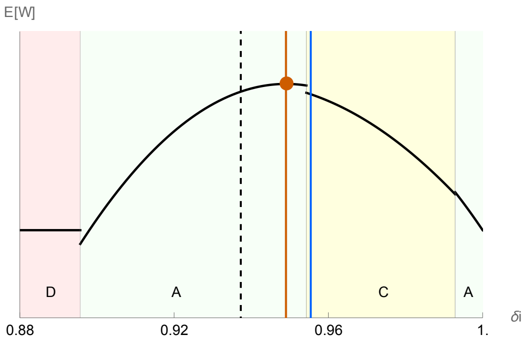
Figure 7d shows an example of high polarisation, that is $\phi > \bar{\phi}^G$, where the challenger never ratifies. Still, the incumbent has an incentive to depart from the no-election treaty parameter: The higher the office rent, the more important it is for the incumbent to increase reelection probability and the more they assimilate towards the median voter. Thus, as R increases, $\delta^*_i \to \delta^*_M$. We can think of this type of treaty as an assimilation treaty.



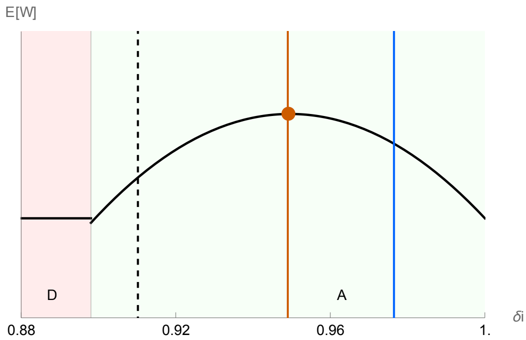
(a) First-best treaty with $\phi = 0.1, R = 1$



(b) Compromise treaty with $\phi = 0.17, R = 1$



(c) Distinction treaty with $\phi = 0.17, R = 1.25$



(d) Assimilation treaty with $\phi = 0.6, R = 1$

Figure 7: Green incumbent treaty outcomes

Numerically, using the introduced parameters, the polarisation threshold values are given by $\bar{\phi}^G_{FB} = 0.134$ and $\bar{\phi}^G = 0.202$. Table 1 details the differences between the compromise and distinction treaty. The threshold office rent $\bar{R}$ that separates a compromise from a distinction treaty roughly corresponds to a relative importance of climate policy of $\omega = 26\%$. If the incumbent draws less than 26% of their welfare from climate policy, they will opt for a distinction treaty, slightly increasing their reelection probability beyond the incumbency advantage. This is also a treaty with higher ambition, which, although better for the median voter than the no-election treaty, is a worse fit than the compromise treaty. Also, because the distinction treaty would only be ratified by the reelected incumbent, the median voter's expected welfare is dampened by the potential non-cooperative outcome of the challenger, which is reflected in the *political fit* measure.

For a higher level of polarisation an assimilation treaty results, as detailed in Table 2. The incumbent optimally chooses a treaty that slightly reduces their reelection probability below the incumbency advantage. This could be prevented by moving closer to the median voter's optimal treaty, however, this is too costly in terms of weakening the treaty since that would mean to forego some emission reductions by country 2.

Table 1: Green incumbent and polarisation $\phi = 0.17$, $\omega(\bar{R} = 1.12) \approx 0.257$

Treaty Type	Emissions (e_i, e_j)		Reelec. Pr.	$R, \omega(R)$	ΔW_M^{TF}	ΔW_M^{PF}
Compromise $\delta_i^* = 0.954 + \epsilon$	$\tilde{e}_1 = 0.898$ $\tilde{e}_2 = 0.907$	$\tilde{e}_1 = 0.898$ $\tilde{e}_2 = 0.907$	$p^C = 0.55$	$R = 1$ $\omega(R) = 0.28$	+0.036%	+0.118%
Distinction $\delta_i^* = 0.949$	$\tilde{e}_1 = 0.834$ $\tilde{e}_2 = 0.902$	$\hat{e}_1 = 0.959$ $\hat{e}_2 = 0.95$	$p^A = 0.5507$	$R = 1.25$ $\omega(R) = 0.24$	+0.032%	+0.017%

Table 2: Green incumbent and polarisation $\phi = 0.6$

Treaty Type	Emissions (e_i, e_j)		Reelec. Pr.	$R, \omega(R)$	ΔW_M^{TF}	ΔW_M^{PF}
Assimilation $\delta_i^* = 0.9496$	$\tilde{e}_1 = 0.874$ $\tilde{e}_2 = 0.902$	$\hat{e}_1 = 0.98$ $\hat{e}_2 = 0.95$	$p^C = 0.5499$	$R = 1$ $\omega(R) = 0.26$	+0.387%	+0.212%

Interestingly, in this highly polarised scenario, election pressure proves beneficial for the median voter. The green incumbent, driven by the assimilation motive, moderates their ambitions, which in the absence of an election are too radical for the median voter. As a result, the median voter's welfare is noticeably higher under an assimilation treaty compared to the outcome of a no-election treaty.

4.4.2 Brown incumbent

When the incumbent party is brown, and the ordering of ratification thresholds follows the structure shown in Figure 4, the reelection probabilities are as outlined in Section 4.3. Once again, we will examine how varying levels of political polarisation influence the optimal treaty choice.

Similarly to the case of a green incumbent, there is a threshold level of polarisation, $\bar{\phi}^B$, beyond which no common ratification interval exists, as defined by (A.17). Additionally, the range in which the no-election treaty is available as a consensus treaty is constrained by $\bar{\phi}^B_{FB}$. The specifics of these two threshold parameters are provided in Lemma A.5. Figure 8 illustrates the two broad polarisation ranges that emerge.

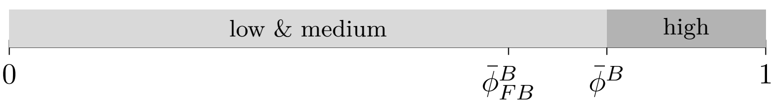


Figure 8: Brown incumbent polarisation ranges

Proposition 5 presents optimal treaty outcomes for a brown incumbent. For a wide range of polarisation levels, consensus treaties are possible. Analogously to the green incumbent, the brown incumbent trades-off the relative importance of policy outcome and chances of reelection and chooses to put more weight on the latter if the office rent is sufficiently high, the threshold value of which is defined in the following:

$$\bar{R}^B(\phi) = \begin{cases} \arg\min |W^C(\hat{\delta}_i) - W^A(\bar{\delta}_j)| & \text{for } \phi \leq \phi^{\bar{B}}_{FB} \\ \arg\min |W^C(\bar{\delta}_j) - W^A(\bar{\delta}_j)| & \text{for } \phi^{\bar{B}}_{FB} < \phi \leq \bar{\phi}^B \\ \arg\min |W^B(\bar{\delta}_j) - W^A(\underline{\delta}_i)| & \text{for } \phi > \bar{\phi}^B. \end{cases} \tag{27}$$

The incumbent will choose a consensus treaty if the office rent lies below (27), and a distinction treaty otherwise. When the first-best treaty is available, the resulting treaties differ in ambition. Otherwise the suggested treaty parameters are only marginally different, but the distinction treaty is only ratified by the incumbent. When polarisation is high, the incumbent chooses an insurance treaty for low levels of the office rent, and a distinction treaty otherwise.

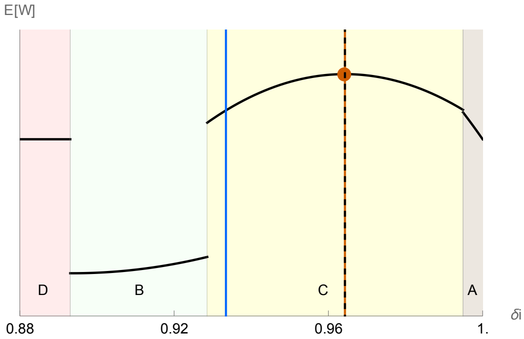
Proposition 5 (Treaty Outcomes for $i = B$)

(i) For low & medium levels of polarisation, that is for $\phi \leq \bar{\phi}^B$, it holds that:

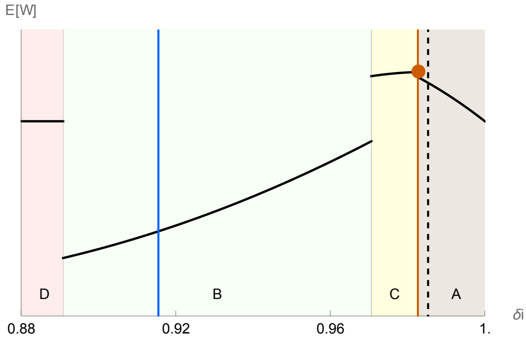
1. *it is never optimal for the brown incumbent to choose a treaty of type B,*
2. *the brown incumbent chooses a consensus treaty if $R \leq \bar{R}^B(\phi)$ and a distinction treaty with $\delta_i^* = \bar{\delta}_j + \epsilon$ otherwise.*
3. *The chosen consensus treaty is first-best for $\phi \leq \bar{\phi}^G_{FB}$, that is $\delta_i^* = \hat{\delta}_i$, otherwise it is a compromise treaty with $\delta_i^* = \bar{\delta}_j - \epsilon$.*

(ii) For high levels of polarisation, that is for $\phi > \bar{\phi}^B$, the brown incumbent chooses an insurance treaty with $\delta_i^* = \bar{\delta}_j$ if $R \leq \bar{R}^B(\phi)$ and a distinction treaty with $\delta_i^* = \underline{\delta}_i$ otherwise.

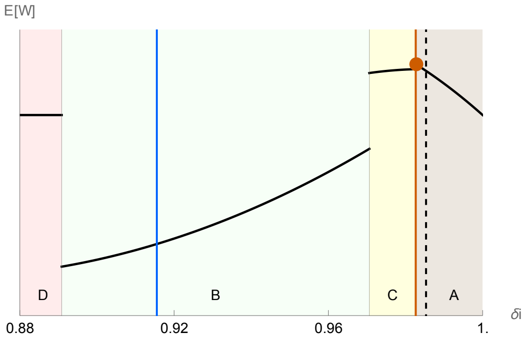
Figure 9 illustrates exemplary treaty outcomes with a brown incumbent. The first-best treaty is available for quite a large range of polarisation levels, an example of which is depicted in 9a. The choice between a compromise and a distinction treaty is illustrated in Figures 9b and 9c, which shows dynamics analogous to the case of a green incumbent, albeit moving in the opposite direction.



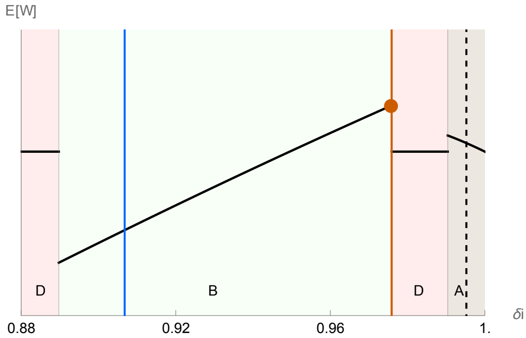
(a) First-best treaty with $\phi = 0.3, R = 1$



(b) Compromise treaty with $\phi = 0.7, R = 1$



(c) Distinction treaty with $\phi = 0.7, R = 1.25$



(d) Insurance treaty with $\phi = 0.9, R = 0.5$

Figure 9: Brown incumbent treaty outcomes

By choosing a distinction treaty, the median voter is forced to compare the weak treaty to the non-cooperative outcome of the highly green challenger, where $\Delta W_M^A > 0$ and thus $p^A(\delta_i^*) > p^C$. The prospect of a high office rent makes it worth for the incumbent to exploit this difference as opposed to choosing a better policy.

Finally, the case of very pronounced polarisation is shown in Figure 9d. In this case, there even exists an interval $(\bar{\delta}_j, \underline{\delta}_i)$ which is not ratified by any of the two parties, as the treaty would be too ambitious for the incumbent and not ambitious enough for the challenger. The choice between an insurance treaty and distinction treaty again depends on the size of the office rent: if R is sufficiently small, it can be optimal for the incumbent to suggest a treaty that they themselves would not ratify but their green challenger would. This strategy is beneficial and “low-risk” for the incumbent for two reasons: on the one hand, they are pursuing a “cheap” policy in case of reelection by not committing to any emission reductions. Given their low environmental preferences, the cost of country 2 not reducing emissions is relatively low. On the other hand, in case they are replaced, the negotiated treaty gives the incumbent higher welfare levels than the challenger’s non-cooperative policy choice would. Interestingly, $\Delta W_M^B(\delta_i^*) < 0$, meaning that the incumbent knowingly reduces their reelection probability. Therefore, as opposed to the differentiation treaty and due to the strong polarisation, they prioritise policy outcome over reelection. One can therefore interpret this choice as an insurance against a potential successor.

For the numerical example the polarisation threshold values are given by $\bar{\phi}_B^{FB} = 0.668$ and $\bar{\phi}^B = 0.79$. Table 3 quantifies the differences between a compromise and a distinction treaty. Compromise is preferable if the relative importance of climate policy is at least 30%. Choosing a distinction treaty leads to a reelection probability above the incumbency advantage, because the challenger's non-cooperative emission choice is costly for the median voter, combined with comparably little emission reductions by country 2. For the median voter, due to the fact that the treaty ambition only marginally differs, the *treaty fit* is identical and a slight improvement over the no-election treaty. However, because the compromise treaty avoids the costly non-cooperative outcome of the green challenger, in terms of *political fit* it is superior to the distinction treaty.

Table 3: Brown incumbent and polarisation $\phi = 0.7$, $\omega(\bar{R} = 1.12) \approx 0.296$

Treaty Type	Emissions (e_i, e_j)		Reelec. Pr.	$R, \omega(R)$	ΔW_M^{TF}	ΔW_M^{PF}
Compromise $\delta_i^* = 0.983 - \epsilon$	$\tilde{e}_1 = 0.968$ $\tilde{e}_2 = 0.936$	$\tilde{e}_1 = 0.968$ $\tilde{e}_2 = 0.936$	$p^C = 0.55$	$R = 1$ $\omega(R) = 0.32$	+0.044%	+0.166%
Distinction $\delta_i^* = 0.983 + \epsilon$	$\tilde{e}_1 = 0.968$ $\tilde{e}_2 = 0.936$	$\hat{e}_1 = 0.915$ $\hat{e}_2 = 0.95$	$p^A = 0.5507$	$R = 1.25$ $\omega(R) = 0.27$	+0.044%	+0.024%

A numerical example for the insurance treaty is given in Table 4. This treaty type only emerges when climate policy has a high relative importance to the incumbent: in this case, they want to prioritise policy outcome over reelection. This is contrasted with a distinction treaty at the same polarisation level. In comparison, median voter welfare is much improved under an insurance treaty: conceptually, this is a way of compromising when no common ratification is possible.

Table 4: Brown incumbent and polarisation $\phi = 0.9$, $\omega(\bar{R} = 0.78) \approx 0.39$

Treaty Type	Emissions (e_i, e_j)		Reelec. Pr.	$R, \omega(\bar{R})$	ΔW_M^{TF}	ΔW_M^{PF}
Insurance $\delta_i^* = 0.9759$	$\hat{e}_1 = 0.995$ $\hat{e}_2 = 0.95$	$\tilde{e}_1 = 0.971$ $\tilde{e}_2 = 0.927$	$p^B = 0.5485$	$R = 0.7$ $\omega(R) = 0.41$	+0.374%	+0.158%
Distinction $\delta_i^* = 0.9904$	$\tilde{e}_1 = 0.971$ $\tilde{e}_2 = 0.927$	$\hat{e}_1 = 0.995$ $\hat{e}_2 = 0.95$	$p^B = 0.5507$	$R = 1$ $\omega(R) = 0.33$	+0.101%	+0.056%

Broadly speaking, from the perspective of the median voter, election pressure has a moderating influence on treaty outcomes across the polarisation spectrum. This moderating effect is somewhat dampened by higher levels of office rent, which incentivises incumbents to seek differentiation. However, this should not obscure the fact that increased polarisation results in less favourable policy outcomes for the median voter in absolute terms – as polarisation rises, policy preferences become increasingly distinct by definition.

Furthermore, this illustration points out two differences between the green and the brown incumbent from the point of view of the median voter: firstly, due to the lower environmental preferences of the brown incumbent, a given treaty parameter translates into lower absolute emission reductions in country 1 compared to a green incumbent with the same treaty. Secondly, the fact that the brown incumbent suggests less strict treaties also implies that country 2 reduces emissions by less, resulting in higher damage externalities than under a green incumbent.

This highlights the fundamental difference between the two governments, best illustrated in Figure 5: the stark contrast in available treaties. For the green incumbent, compared to a no-election scenario, moderation results in lower emission reductions abroad but higher domestic costs due to lower status quo emissions. Conversely, a compromise by the brown incumbent leads to higher treaty ambition than in the absence of an election, resulting in greater emission reductions abroad.

5 Extensions

While the basic model framework effectively captures the core dynamics of the issue, it lends itself to a number of extensions that could further refine the results. Each extension introduces additional nuances, offering a more detailed and comprehensive understanding of the model dynamics.

5.1 Treaty Emissions as an Upper Bound

While I assume that instant renegotiation of a treaty is not possible, it could be argued that a government in power can always go beyond the promises made in an international treaty. Presumably, country 2 would not oppose to country 1 reducing emissions by more than what was agreed upon. While I would argue that this case is hardly seen empirically, illustrated by the lack of countries which overshoot their emission pledges, allowing for the elected government to go beyond treaty targets affects some outcomes of the model in an interesting fashion. In this extension, we will therefore interpret treaty emissions $\tilde{e}_1$ as an upper bound which the elected government can voluntarily undercut.

First, note that this change in assumption does not affect the equilibrium outcomes in the case of a brown challenger since the brown party's non-cooperative emission choice is never lower than treaty emissions negotiated by the green incumbent:

$$\underbrace{1 - \beta\theta_{j=B}}_{\hat{e}_{j=B}} < \underbrace{\delta_{i=G}(1 - \beta\theta_{i=G})}_{\tilde{e}_{i=G}}, \qquad \text{since} \;\; \theta_{i=G} \geq \theta_{j=B} \;\; \text{and} \;\; \delta_i \in [0,1].$$

Therefore, given that the brown challenger optimally wants to set higher emissions than the treaty emissions and due to the concavity of their welfare function, under a ratified treaty they cannot do better by choosing $e < \tilde{e}_{i=G}$.

Yet, it is possible for the green challenger to have lower non-cooperative emissions compared to treaty emissions negotiated by the brown incumbent, as previously illustrated by Figure 2. More precisely, this is the case whenever the treaty parameter is above a lower bound level δ^{LB}:

$$\begin{aligned} &\underbrace{1 - \beta\theta_{j=G}}_{\hat{e}_{j=G}} \leq \underbrace{\delta_{i=B}(1 - \beta\theta_{i=B})}_{\tilde{e}_{i=B}} \\ &\qquad \Rightarrow \delta_i \geq \delta^{\text{LB}} \equiv \frac{1 - \beta(1+\phi)}{1 - \beta(1-\phi)}. \end{aligned} \tag{28}$$

Note that $\frac{d\delta^{\text{LB}}}{d\phi} < 0$, that is the more polarised the parties are, the lower is this lower bound.

Therefore, if the non-cooperative emission level is feasible ($\hat{e}_j \leq \tilde{e}_{1,i}$), which is the case whenever $\delta_{i=B} \geq \delta^{\text{LB}}$, the elected challenger will ratify the treaty and then choose $\hat{e}_j$. This makes sense intuitively: they cannot do better than to choose their individually optimal emission level, while at the same time getting the treaty benefit of country 2 reducing their emissions below their non-cooperative level, resulting in lower damage costs. Due to the assumption of a linear damage function, emission levels are dominant strategies and a lower than agreed emission level of country 1 does not affect the emission choice in country 2.

This now gives rise to two new cases C' and B', which emerge depending on the level of polarisation as illustrated in Figure 10 and are separated by the polarisation threshold level $\bar{\phi}^{LB}$, that is, where $\underline{\delta}_i$ and δ^{LB} are equal:

$$\begin{aligned} &\underline{\delta}_i = \delta^{\text{LB}} \\ \Rightarrow \;\; &\bar{\phi}^{LB} = \frac{\beta - 1 + \sqrt{1-\beta}}{\beta} \end{aligned} \tag{29}$$

If polarisation is below $\bar{\phi}^{LB}$ it holds that $\underline{\delta}_i < \delta^{\text{LB}}$. This means that there exists a common ratification interval (formerly case C), however, in which the green challenger chooses non-cooperative emissions after ratification. This new case C' thus indicates a range where, depending on who wins the election, i would ratify and set treaty emissions and j would ratify and choose $\hat{e}_j$.

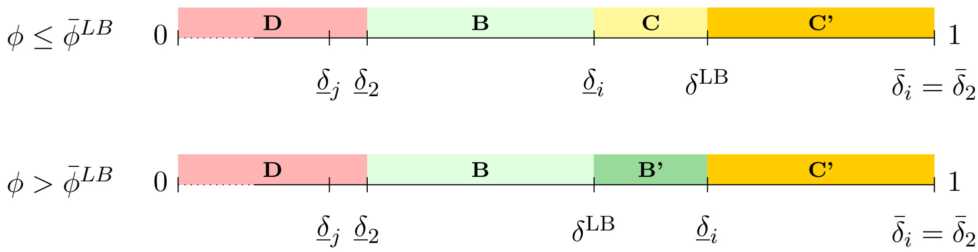
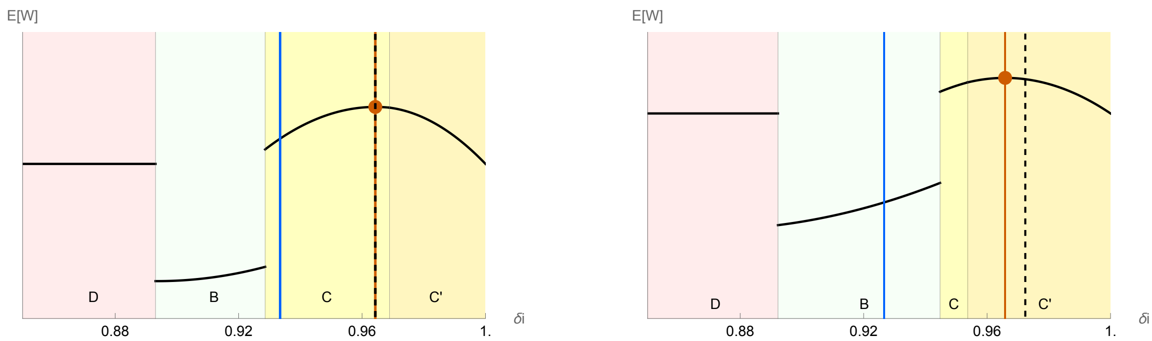


Figure 10: New cases depending on polarisation levels

Polarisation above $\bar{\phi}^{LB}$ leads to the fact that $\underline{\delta}_i > \delta^{\mathrm{LB}}$ and therefore there exists a range where only the green challenger ratifies (formerly case B) and then opts to choose non-cooperative emissions. In this new case B', depending on who wins the election, i would thus not ratify and set non-cooperative emissions $\hat{e}_i$ and j would ratify and choose $\hat{e}_j$. Note that this outcome differs from case D in that country 2 will set treaty emissions.

Independent of polarisation levels, we find that classic distinction treaties no longer exist. This is intuitive: the only way for a brown incumbent to differentiate from a green challenger in the basic model was to negotiate a treaty too weak for the challenger to ratify. Now, however, the challenger ratifies any treaty $\delta \in [\underline{\delta}_j, 1]$. The compromise and insurance treaty types still exist, albeit resulting from slightly different motives.



(a) First-best treaty with $\phi = 0.3, R = 1$

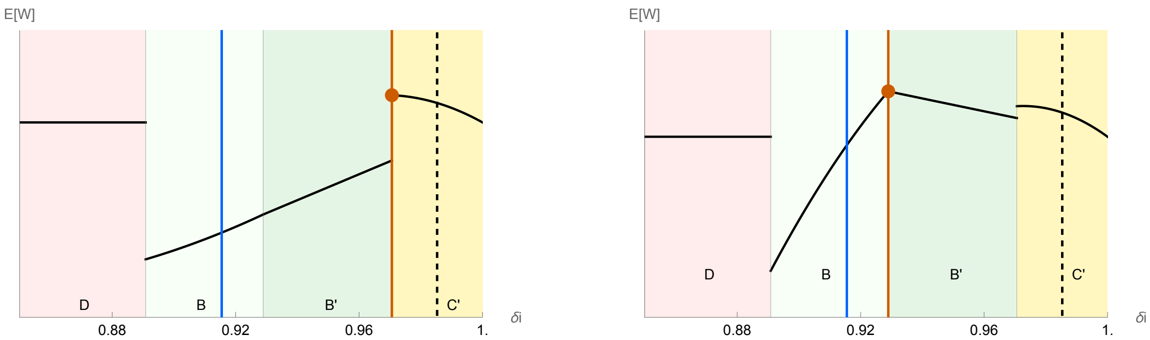
(b) "Compromise" treaty with $\phi = 0.45, R = 1$

Figure 11: Treaty outcomes with low polarisation

With polarisation levels below $\bar{\phi}^{LB}$, which here corresponds to $\bar{\phi}^{LB}(0.05) = 0.494$, first-best and a variation of a compromise treaty are possible as seen in Figure 11. The former occurs when polarisation is sufficiently low. The latter differs from the original compromise treaty in the sense that we find ourselves in the range of case C', where even though the challenger ratifies, they choose non-cooperative emissions. Still, country 2 engages in emission reductions as defined by the treaty. Note that even though $\hat{\delta}_i$ is available within case C', the incumbent optimally chooses a slightly more ambitious treaty. This is due to the fact that in case of election loss, the challenger will in any case choose non-cooperative emissions, however, country 2 will engage in more emission reductions if the treaty parameter is lower, which compensates for a slightly lower reelection probability.

In the case of high polarisation, that is for $\phi > \bar{\phi}^{LB}$, another difference to the main model materialises. Now the green challenger cannot be "locked" in with a weak treaty as before, since they can go beyond treaty emission reductions and thus the classic insurance motive is no longer available for the brown incumbent.

Analogously to the low polarisation case and for standard office rent levels, we observe a type of compromise treaty, an example of which is shown in Figure 12a. We are in area C', where both parties ratify the treaty, and where the incumbent opts for a treaty which is stricter than the no-election treaty in order to achieve higher emission reductions by country 2. Now interestingly, as the office rent decreases and thus less importance is put on securing an election victory, a new variation of an insurance treaty emerges, as illustrated in Figure 12b.



(a) "Compromise" treaty with $\phi = 0.7, R = 1$

(b) "Insurance" treaty with $\phi = 0.7, R = 0.05$

Figure 12: Treaty outcomes with high polarisation

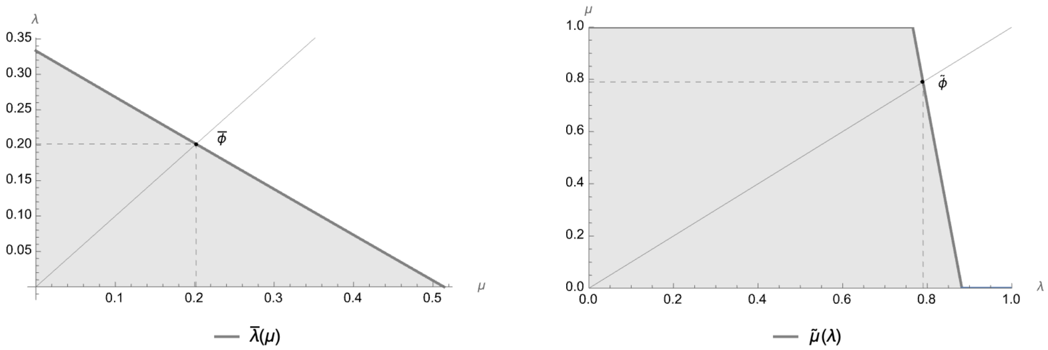
The darker green range indicates case B', where only the challenger ratifies the agreement but for any treaty parameter chooses non-cooperative emissions. In the basic model, the incumbent chose the upper limit of this range to lock in a cheap treaty because the challenger was bound to the treaty emissions, while now, they optimally propose the other end of the range δ^{LB}. At this point, treaty emissions exactly equal the green challenger's non-cooperative emissions. Intuitively, this is optimal because at any other point of the dark green range, the challenger would also set non-cooperative emissions, while this is the point at which emission reductions by country 2 are maximised. Unchanged to the basic model, in case of an election win, the incumbent does not ratify the treaty. So conceptually, this treaty choice still resembles an insurance treaty in the sense that the incumbent aims at optimising the outcome in case of election loss.

5.2 Preference Asymmetry

The basic model assumes the parties' preference parameters to be symmetric around the median voter. Allowing for preference asymmetry, that is for preference parameters that differ in terms of distance to the median voter, not only reproduces all of the presented results, but produces even more distorted outcomes. In particular, we define preference parameters to be:

$$\theta_{1,G} = 1 + \mu, \qquad \theta_{1,B} = 1 - \lambda.$$

The degree of preference asymmetry is captured by the ratio of μ and λ, the symmetric cases being defined by $\mu = \lambda$. Emission choices still follow as described in Section 4.1. There are a few changes in the ratification stage from the results described in Section 4.2. Firstly, in the case of a green incumbent, the threshold value in Lemma A.1 now becomes two-dimensional, as pictured in Figure 13a: Any combination of μ and λ within the shaded area ensures that the brown challenger's ratification interval exists, the boundary of which is the function $\bar{\lambda}(\mu)$. Similarly for the brown incumbent, the threshold value to separate the two scenarios in Figure 4, as defined by Lemma A.2, also becomes two-dimensional as illustrated in Figure 13b. The shaded area then indicates the parameter combinations for which an overlap in ratification intervals exists, the boundary being the function $\tilde{\mu}(\lambda)$.



(a) Threshold value analogy to $\bar{\phi}^G$

(b) Threshold value analogy to $\bar{\phi}^B$

Figure 13: Two-dimensional analogy to threshold values $\bar{\phi}^G$ and $\bar{\phi}^B$ for $\beta = 0.05$

For a more detailed and formal discussion of the changes in the ratification stage, refer to Appendix C. The election stage is qualitatively unaffected by the introduction of preference asymmetry. In the agreement stage, all of the outcomes presented in the symmetry case can also be replicated with asymmetry.

5.3 More sophisticated country 2

It could be argued that in the assumptions of the basic model, country 2 acts in an overly naive fashion, ignoring the possibility of a potential government change in country 1 while taking their participation decision. Here I will relax this assumption by allowing for a slight sophistication in country 2: having an understanding of the political environment in country 1, country 2 can anticipate reelection probabilities prior to negotiations captured by the incumbency advantage z. We will see that this does not affect treaty outcomes for the green incumbent as discussed in Section 4.4, while it reduces the options of a brown incumbent in some instances.

This relaxation of assumption means that country 2 will only accept the suggested treaty if their *expected* welfare change is non-negative, therefore affecting the participation condition (13). Given a treaty parameter δ_i, country 2 can anticipate country 1's ratification decision conditioned on which party will be elected, and therefore compute an expected welfare level as follows:

$$\mathbb{E}[W_2|\delta_i] = z\Big[W_2(\text{'i in power'})\Big] + (1-z)\Big[W_2(\text{'j in power'})\Big] \tag{30}$$

Thus, country 2 will participate in an agreement if their expected welfare is non-negative, which depends on the case resulting from the treaty parameter, as given by the following:

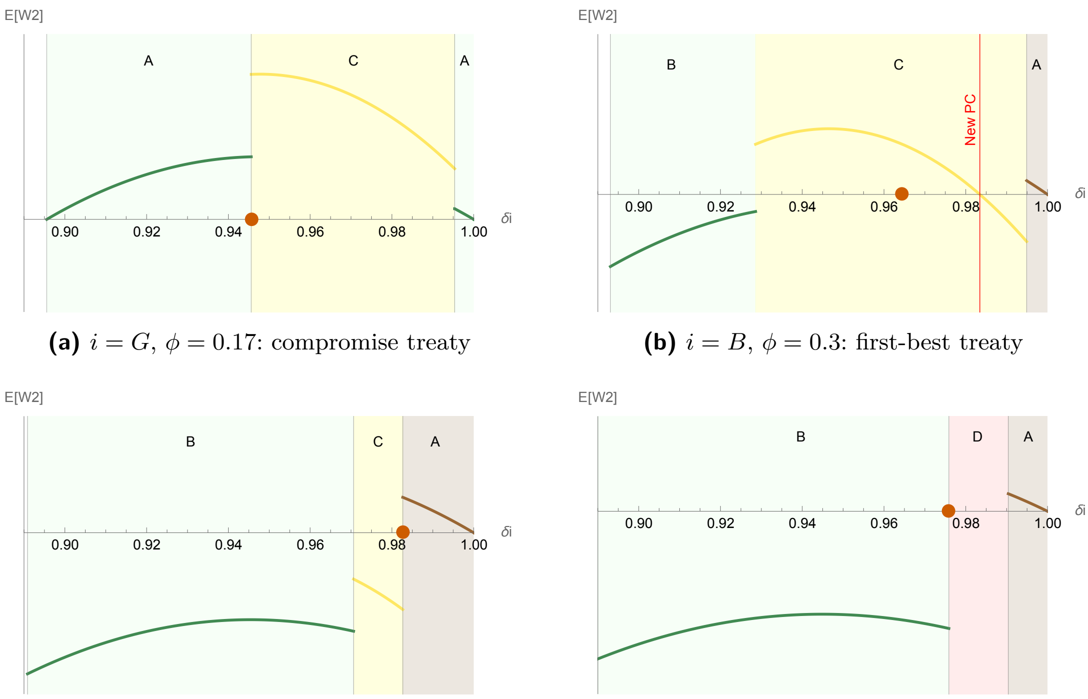
$$\begin{aligned}\Delta\mathbb{E}[W_2^A] &= z\Big[\tilde{W}_2(\delta_i,\theta_i) - \hat{W}_2(\theta_i)\Big] + (1-z)\Big[\overbrace{\hat{W}_2(\theta_j) - \hat{W}_2(\theta_j)}^{=0}\Big] \\ &= z\Big[\tilde{W}_2(\delta_i,\theta_i) - \hat{W}_2(\theta_i)\Big] \end{aligned} \tag{31a}$$

$$\begin{aligned}\Delta\mathbb{E}[W_2^B] &= z\Big[\overbrace{\hat{W}_2(\theta_i) - \hat{W}_2(\theta_i)}^{=0}\Big] + (1-z)\Big[\tilde{W}_2(\delta_i,\theta_i) - \hat{W}_2(\theta_j)\Big] \\ &= (1-z)\Big[\tilde{W}_2(\delta_i,\theta_i) - \hat{W}_2(\theta_j)\Big] \end{aligned} \tag{31b}$$

$$\Delta\mathbb{E}[W_2^C] = \tilde{W}_2(\delta_i,\theta_i) - \Big[z\hat{W}_2(\theta_i) + (1-z)\hat{W}_2(\theta_j)\Big] \tag{31c}$$

We will now analyse whether this changed participation constraint will affect any of the equilibrium outcomes in Section 4.4. First note that (31a) is qualitatively unchanged to the original participation constraint, since only the incumbent ever ratifies and therefore, any treaty of case A, that is all distinction and assimilation treaties, will still be valid.

Figure 14 displays all consensus and insurance treaties from Section 4.4. The shaded areas indicate the treaty type which arises in a specific range, as a consequence of ratification intervals in country 1. The plotted functions follow (30) for the respective cases and in corresponding colours (numerical example with $\beta = 0.05, z = 0.55$). Note that this differs from functions plotted in Figures 7 and 9, where the functions in black represent the country 1 incumbent's expected welfare, which determines the optimal treaty choice. Here, whenever the function takes a non-negative value, country 2 participates in the treaty.



(a) $i = G$, $\phi = 0.17$: compromise treaty

(b) $i = B$, $\phi = 0.3$: first-best treaty

(c) $i = B$, $\phi = 0.7$: compromise treaty

(d) $i = B$, $\phi = 0.9$: insurance treaty

Figure 14: Expected welfare of country 2 and optimal treaty choice for incumbent

For a green incumbent, the results are unaffected: any distinction and assimilation treaty is still valid and, as pictured in Figure 14a, the proposed compromise treaty yields a positive expected welfare for country 2. However, the new participation constraints of country 2 affect the findings for a brown incumbent: Figure 14b illustrate the case of low polarisation, where a first-best treaty is proposed. The new participation constraint reduces the range of feasible consensus treaties, yet, if polarisation is sufficiently low (detailed in Lemma C.7), the new constraint (red line) always lies above the optimal treaty choice (orange dot) for low polarisation values and therefore does not affect equilibrium outcomes. This does not hold true for higher levels of polarisation: the compromise treaty (Figure 14c) and the insurance treaty (Figure 14d) are no longer an option for the brown incumbent, being ruled out by the participation constraint of a the more sophisticated country 2.

The intuition of why the insurance treaty collapses is straightforward: if country 2 anticipates a government change in country 1, they would prefer green non-cooperative emissions and no emission reduction costs on their side over a weak treaty. In a similar vein, with medium polarisation, they refuse a weak compromise treaty in favour of the non-cooperative outcome, since they anticipate the low non-cooperative emissions by the green challenger as an option, combined with no emission reduction costs. In essence, a more sophisticated country 2 approaches ratification decisions with strategic foresight, speculating on a possible government change. This allows them to counteract some of the distortionary incentives of a brown incumbent in country 1. However, this sophistication also comes with a side-effect: at medium levels of polarisation, it prevents the adoption of a compromise treaty more favourable to the median voter, while still allowing for the more harmful distinction treaty.

6 Discussion & Conclusions

The interaction between political economy and international environmental cooperation has received comparably little attention in the literature so far. A deeper understanding of major political economy frictions is thus paramount for the creation of more successful treaties and policies in the future. This paper speaks to the importance of considering political polarisation in the context of domestic elections as a crucial element in international environmental policy.

To explore this issue, I examine how domestic elections influence the formation of IEAs and assess whether they help explain the limited success of current global cooperation on climate change mitigation. Decision-makers in international negotiations often face electoral pressures at home, meaning that their policy choices can impact their prospects in future elections. This dynamic, in turn, influences their choice of optimal policy. Furthermore, elections introduce the possibility that one government may negotiate an agreement while another may handle its ratification, a factor frequently overlooked in the existing literature.

I find that incumbent governments have clear incentives to adjust their policy choices to enhance their prospects in upcoming elections. This dynamic leads to two distinct types of treaties: *consensus* and *differentiation*. In the case of consensus treaties, incumbents place high value on the policy outcome itself, choosing to accommodate the preferences of the opposing party in order to ensure ratification, even if they lose the election. Differentiation treaties, on the other hand, arise when reelection takes precedence. Here, the incumbent seeks to distance themselves from their opponent, forcing voters to choose between two distinct environmental policy paths. The incumbent's objective function reflects this trade-off: while choosing the treaty that maximises their in-office welfare, that is $\hat{\delta}_i$, might seem optimal in isolation, it may not be the best choice when considering its impact on reelection chances. It can therefore be strategically advantageous for incumbents to adjust treaty parameters such as to improve reelection prospects, but also to secure a more favourable welfare outcome in the event of an electoral loss. The incentive to prioritise reelection becomes particularly strong when the office rent is high, reducing the relative importance of policy outcomes and making the desire to remain in power paramount.

Throughout the paper, I examine the cases of a *green* and a *brown* incumbent. Both share the characteristic that when polarisation is sufficiently low, consensus, and even the incumbent's first-best outcome may be achieved, leading to mutual ratification of treaties. However, as polarisation intensifies, consensus becomes infeasible, and more polarised outcomes emerge. A key distinction between the two incumbents lies here: while a stringent treaty under a green incumbent imposes high emission reduction costs on country 1, it also results in significant reductions in country 2 and thus lower damage costs. Conversely, in a highly polarised environment, brown incumbents either choose not to ratify or favour shallow treaties. Due to their weaker environmental preferences, this leads to high emissions in country 1 and thus low emission reduction costs, combined with minimal reduction efforts in country 2. In summary, although the median voter is positioned symmetrically between the two parties, the resulting welfare losses under each government are asymmetrical.

The presented model also allows us to observe the distinct policy dilemmas faced by the two parties. A green party with ambitious environmental goals is invariably forced to scale back its agenda. If it fails to do so, it risks undermining its reelection prospects, with the significant risk of leading to the brown party's non-cooperative outcome, which is more detrimental than the status quo. Moreover, the green party faces a narrow window of polarisation levels that allow for consensus in the first place; beyond this, they must appeal to the median voter by reducing ambition in order to avoid losing the election. In any case, if the green party were to stick to its platform, the likelihood of being replaced becomes considerable, potentially not only stalling progress on climate change mitigation but also reversing it. The brown party operates from a very different baseline. Across a wide range of polarisation levels, their optimal strategy is to fully implement their platform, resulting in consensus treaties when office rents are sufficiently low – an outcome that also improves on the welfare of the median voter. Even under high polarisation, the brown incumbent has a low-risk strategy: they may choose not to ratify the treaty but can still limit its stringency if they lose the election. In either case, the result is minimal or no progress on climate change mitigation in both countries.

The model setup includes a number of exogenous political economy factors: the office rent, the incumbency advantage and the density of the popularity shock all capture specific characteristics of a political system and therefore would allow for a discussion of the effects of domestic elections on international cooperation for a wide array of political landscapes. The size of the office rent determines the weight that the incumbent puts on the maximisation of the reelection probability as opposed to achieving a favourable policy outcome. Similarly to Battaglini and Harstad (2020) this turns out to be the deciding factor between consensus and differentiation treaties. However, increased polarisation leads to the fact that consensus treaties are no longer available and that new scenarios emerge: a green incumbent chooses to assimilate towards the median voter, therefore weakening the treaty independent of office rent levels, while the brown incumbent opts for an insurance treaty, not committing to any emission reductions in case of reelection and handing a weak treaty to their challenger in case of election defeat. Both of these results are novel and underline the importance of accounting for political polarisation.

Although increasing polarisation generally reduces the absolute welfare of the median voter – simply because party preferences are further removed from median voter preferences – I find that electoral pressure can lead to outcomes more favourable to the median voter than those arising in the absence of elections. This moderating effect of elections becomes more pronounced as polarisation increases, as elections often help to prevent the extreme outcomes that would otherwise occur without electoral accountability.

I propose several extensions to further explore key dynamics of the model. By allowing the election winner to freely undercut treaty emissions, I demonstrate that under a brown incumbent, no treaty is too weak for the green challenger to ratify, as they can voluntarily commit to lower emissions. This effectively eliminates the option of a distinction treaty from the incumbent's set of strategies. Similarly, if country 2 is sophisticated enough to anticipate a potential change in government from their treaty partner, it further limits the brown incumbent's options. At higher levels of polarisation, country 2 would prefer non-cooperative green emissions over being drawn into a weak treaty by country 1. However, this can sometimes negatively impact the median voter, as favourable consensus treaties may be taken off the table.

A clear limitation of this model is its focus on a two-party democracy and a bilateral treaty, whereas effective climate mitigation will require multilateral cooperation. Future research could explore how the presence of additional parties across the political spectrum might mitigate the effects of polarisation. While the reduction to a two-country scenario is chosen for analytical clarity, meaningful international climate policy indeed ultimately depends on a few key global players. In this context, the model remains valuable if we consider country 1 as a pivotal large actor and country 2 as representing the rest of the world, willing to engage in an environmental agreement.

Finally, the relevance of this model extends beyond its specific context. In general, whenever cross-border public goods are involved and governance is divided among multiple bodies, the policy versus election trade-off described here will likely arise. It follows that in most cases, the underprovision of public goods is exacerbated by the interplay between local electoral pressures and heightened political polarisation. This mirrors previous findings in the literature, for example on country's reduced willingness to invest in international cooperation, for example international security, with increased levels of polarisation (Esteban and Schneider 2008; Perrings et al. 2021). Further research is needed to translate these insights into, for example, treaty mechanisms or institutional frameworks that can leverage these incentive structures to ultimately achieve more efficient outcomes.

Appendix

A Additional Results

Proposition A.1 (Stage 3: Ratification Intervals with $i = G$)
In the case of a green incumbent, the incumbent's and country 2's ratification thresholds are given as follows:

$$[\underline{\delta}^{i=G}, \bar{\delta}^{i=G}] = \left[\max\left\{0, \frac{1+\beta(1+\phi)\left[\beta(3+\phi)-4\right]}{(\beta(1+\phi)-1)^2}\right\},\ 1\right], \tag{A.1}$$

$$[\underline{\delta}_2(\theta_G), \bar{\delta}_2(\theta_G)] = \left[\frac{1+\beta[\beta(3+2\phi)-4]}{(\beta-1)^2},\ 1\right]. \tag{A.2}$$

The challenger's ratification thresholds exist when $\phi \leq \bar{\phi}^G$. In that case, they are given by:

$$[\underline{\delta}^{j=B}, \bar{\delta}^{j=B}] = \left[\frac{1+\beta\left[\phi-3\right]-\beta^2\left[\phi^2+\phi-2\right]-\sqrt{M_{j=B}}}{(\beta(1+\phi)-1)^2},\right.$$
$$\left.\frac{1+\beta\left[\phi-3\right]-\beta^2\left[\phi^2+\phi-2\right]+\sqrt{M_{j=B}}}{(\beta(1+\phi)-1)^2}\right], \tag{A.3}$$

where $M_{j=B} = \beta^2(\beta-1)(\phi-1)\left[1-5\phi+\beta(4\phi^2+5\phi-1)\right]$.

Proof of Proposition A.1
(i) Ratification interval for $i = G$:
Threshold values follow from (9), which is a concave function and has two roots. The upper ratification threshold is equal to 1 because it corresponds to the incumbent's non-cooperative emission choice and thus makes them equally well off as without a treaty.

To show when $\underline{\delta}^{i=G}$ is non-negative, consider the value of $\phi \in [0,1]$ that renders $\underline{\delta}^{i=G} = 0$:

$$\underline{\delta}^{i=G}(\phi_0^{i=G}) = 0$$
$$\Rightarrow\ \phi_0^{i=G} = \frac{2-\beta-\sqrt{3-4\beta+\beta^2}}{\beta}\ .$$

Knowing that the lower ratification threshold for the green incumbent decreases in ϕ (see Proposition A.2) we can then show that ϕ_0 is only restrictive if marginal damages are sufficiently high:

$$\phi_0 \leq 1\ \Rightarrow\ \beta \leq 0.1465.$$

Consequently, as long as $\beta \leq 0.1465$, it holds that $\underline{\delta}^{i=G} \geq 0$.

(ii) Ratification interval for country 2:
Threshold values follow from (13), which is a concave function and has two roots.

The lower ratification threshold is always positive:

$$\underline{\delta}_2 = \frac{1+\beta[\beta(3+2\phi)-4]}{(\beta-1)^2} > 0$$

$$\frac{1-4\beta+\overbrace{\beta^2(3+2\phi)}^{\geq 0}}{\underbrace{(\beta-1)^2}_{>0}} > 0$$

$$\Rightarrow \quad 1-4\beta > 0 \qquad \forall \beta \in [0, 0.15]\ .$$

The upper ratification threshold is equal to 1 because it corresponds to their non-cooperative emission choice and thus makes them equally well off as without a treaty.

(iii) Ratification interval for $j = B$:
Threshold values follow from (9), which is a concave function and has two roots. I will now show that the lower ratification threshold is always positive. Since the denominator is quadratic, it suffices to look at the sign of the numerator:

$$1+\beta\left[\phi-3\right]-\beta^2\left[\phi^2+\phi-2\right]-\sqrt{M_{j=B}} \lesseqgtr 0$$

Note that ratification thresholds for the brown challenger only exist if $\phi \leq \bar{\phi}^G$, where $\bar{\phi}^G$ is maximised at $\beta = 0.15$ and $\bar{\phi}^G(0.15) \approx 0.206$. Also, $M_{j=B}$ is maximised at $\beta = 0.15$ and $\phi = 0$ and at those parameter values is $M_{j=B} \approx 0.02$, and consequently $\sqrt{M_{j=B}} \approx 0.14$. Therefore:

$$\underbrace{1+\beta\overbrace{\left[\phi-3\right]}^{\in[-3,-2.794]}-\beta^2\overbrace{\left[\phi^2+\phi-2\right]}^{\in[-2,-1.752]}}_{\in[0.589,0.6259]} > \underbrace{\sqrt{M_{j=B}}}_{\in[0,0.14]}$$

Since the two ranges do not overlap, the numerator and hence the lower ratification threshold of the brown incumbent is always positive.

Next, we will show that the upper ratification threshold is lower than 1. Restating it as follows:

$$\frac{\mathcal{A}+\sqrt{M_{j=B}}}{\mathcal{B}} < 1,$$

and then reformulating:

$$\begin{aligned}\sqrt{M_{j=B}} < \mathcal{B}-\mathcal{A} \equiv \mathcal{C} \;\Leftrightarrow\; & \mathcal{C}^2 - M_{j=B} > 0 \\ \Rightarrow\; & 4\beta^2\phi^2(\beta(1+\phi)-1)^2 > 0 \qquad \text{which is true.}\end{aligned}$$

Therefore, we have shown that $\frac{\mathcal{A}+\sqrt{M_{j=B}}}{\mathcal{B}} < 1$. □

Lemma A.1 (Existence of Ratification Interval for Brown Challenger)
The challenger's ratification thresholds exist if it holds that:

$$\phi \leq \bar{\phi}^G(\beta) = \frac{5\beta - 5 + \sqrt{(\beta-1)(41\beta-25)}}{8\beta}. \tag{A.4}$$

In addition it holds that:

$$\frac{d\bar{\phi}^G}{d\beta} > 0. \tag{A.5}$$

Proof of Lemma A.1
(i) The challenger's threshold values exist if the term in the square root in (A.3), that is, $M_{j=B}$ is non-negative. Therefore:

$$\underbrace{\beta^2\overbrace{(\beta-1)}^{<0}\overbrace{(\phi-1)}^{<0}}_{>0}\Big[1-5\phi+\beta(4\phi^2+5\phi-1)\Big] \geq 0\ .$$

It thus suffices to consider the term in square brackets to determine the sign:

$$\begin{aligned}&1-5\phi+\beta(4\phi^2+5\phi-1) \geq 0 \\ &\qquad \phi \leq \frac{5-5\beta+\sqrt{(\beta-1)(41\beta-25)}}{8\beta} \equiv \bar{\phi}^G\ .\end{aligned}$$

(ii) Proof of (A.5):

$$\frac{d\bar{\phi}^G}{d\beta} = \frac{25-33\beta-5\sqrt{(\beta-1)(41\beta-25)}}{8\beta^2\sqrt{(\beta-1)(41\beta-25)}}$$

The term under the square root is non-negative since $\beta - 1 < 0$ and $41\beta - 25 < 0$ $\forall \beta \in [0, 0.15]$. The sign is thus determined by the numerator:

$$
\begin{aligned}
25 - 33\beta - 5\sqrt{(\beta-1)(41\beta-25)} &\lesseqgtr 0 \\
(25-33\beta)^2 &\lesseqgtr 25(\beta-1)(41\beta-25) \\
625 + 1089\beta^2 - 1650\beta &\lesseqgtr 625 + 1025\beta^2 - 1650\beta \\
64\beta^2 &> 0 \ ,
\end{aligned}
$$

and therefore $\frac{d\bar{\phi}^G}{d\beta} > 0$.

□

Proposition A.2 (Stage 3: Comparative Statics with $i = G$)
The following conditions hold for the equilibrium ratification intervals under the condition that thresholds exist and that they are within the interval $[0, 1]$*:*

$$\frac{d\underline{\delta}^{i=G}}{d\phi} < 0, \qquad \frac{d\bar{\delta}^{i=G}}{d\phi} = 0, \qquad \frac{d\underline{\delta}^{j=B}}{d\phi} > 0, \qquad \frac{d\bar{\delta}^{j=B}}{d\phi} < 0.$$

Proof of Proposition A.2
(i) For the green incumbent's thresholds:

$$\frac{d\underline{\delta}^{i=G}}{d\phi} = \frac{-2\beta \overbrace{(\beta-1)}^{<0} \overbrace{(1+\beta(1+\phi))}^{>0}}{\underbrace{(\beta(1+\phi)-1)^3}_{<0}} < 0 \tag{A.6}$$

The upper ratification threshold of the green incumbent is a constant and therefore not a function of ϕ.

(ii) For the brown challenger's thresholds:

- $\frac{d\underline{\delta}^{j=B}}{d\phi} > 0$

$$\frac{d\underline{\delta}^{j=B}}{d\phi} = \frac{\beta(\beta-1)\Big[\beta^2(\phi^2 - 15\phi - 8) + \beta^3(2\phi^3 - 5\phi^2 - 10\phi + 5) + 3\sqrt{M_{j=B}} -}{(\beta(1+\phi)-1)^3\sqrt{M_{j=B}}}$$
$$\frac{\beta(-3 + 5\sqrt{M_{j=B}} + \phi(5 + \sqrt{M_{j=B}}))\Big]}{(\beta(1+\phi)-1)^3\sqrt{M_{j=B}}} \tag{A.7}$$

I will now show that this term is positive. The denominator is negative. In the numerator, because $\beta(\beta-1) < 0$, I will show that the term in square brackets is positive. Rewriting this term:

$$\sqrt{M_{j=B}}(\underbrace{3-5\beta-\beta\phi}_{\mathcal{A}}) + \phi^3(\underbrace{2\beta^3}_{>0}) + \phi^2(\underbrace{\beta^2-5\beta^3}_{\mathcal{B}}) + \phi(\underbrace{15\beta^2-10\beta^3}_{>0}) \\ +\beta(\underbrace{3-5\phi+5\beta^2-8\beta}_{\mathcal{C}}) \tag{A.8}$$

where

$$\mathcal{A} = 3-5\beta-\beta\phi > 0 \quad \text{for } \beta < \frac{3}{5-\phi} \in [0.58, 0.6], \text{ depending on } \phi \in [0, \bar{\phi}^B].$$
$$\mathcal{B} = \beta^2 - 5\beta^3 > 0 \quad \text{for } \beta < \frac{1}{5}$$
$$\mathcal{C} = 3-5\phi+5\beta^2-8\beta > 0 \text{ (see below)}$$

Note that the expression $\mathcal{C}$ is strictest at $\beta = 0.15$ and $\phi = \bar{\phi}^G \approx 0.206$, where $\mathcal{C}(\beta = 0.15, \phi = 0.206) = 0.7925 > 0$. Consequently, since all terms in (A.8) are positive, the expression in square brackets is also positive. Hence the numerator is negative, and combined with the negative denominator, it holds that $\frac{d\underline{\delta}^{j=B}}{d\phi} > 0$.

- $\frac{d\bar{\delta}^{j=B}}{d\phi} < 0$

$$\frac{d\bar{\delta}^{j=B}}{d\phi} = \frac{\beta(1-\beta)\Big[\beta^2(\phi^2+15\phi-8)+\beta^3(2\phi^3-5\phi^2-10\phi+5)-3\sqrt{M_{j=B}}+}{(\beta(1+\phi)-1)^3\sqrt{M_{j=B}}} \\ \frac{\beta(3+5\sqrt{M_{j=B}}+\phi(-5+\sqrt{M_{j=B}}))\Big]}{(\beta(1+\phi)-1)^3\sqrt{M_{j=B}}} \tag{A.9}$$

I will now show that this term is negative. Again, the denominator is negative, so that it remains to be shown that the numerator is positive. First, $\beta(1-\beta) > 0$. Then, rewriting the numerator:

$$\mathcal{A} + [-3+5\beta+\beta\phi]\sqrt{M_{j=B}} > 0$$
$$\mathcal{A} > \underbrace{[3-5\beta-\beta\phi]}_{>0}\sqrt{M_{j=B}}$$
$$\frac{\mathcal{A}^2}{[3-5\beta-\beta\phi]^2} > M_{j=B}$$

Therefore:

$$M_{j=B} - \left(\frac{\mathcal{A}}{[-3+5\beta+\beta\phi]}\right)^2 < 0$$

which expands to

$$\frac{1}{(\cdot)^2}\left(4\beta^2\phi\underbrace{(\beta(\phi-1)-1)^3}_{<0}\underbrace{[6-\overbrace{5\phi}^{\max=1.03}-\overbrace{\beta(10-9\phi+\phi^2)}^{\max=1.5}]}_{>0}\right) < 0.$$

Consequently, the term in square brackets in the numerator is positive, making the whole expression negative.

□

Proposition A.3 (Stage 3: Ratification Intervals with $i = B$)
In the case of a brown incumbent, ratification thresholds are given as follows:

$$[\underline{\delta}^{i=B}, \bar{\delta}^{i=B}] = \left[\frac{1+\beta(\phi-1)\,[4+\beta(\phi-3)]}{(1+\beta(\phi-1))^2}, 1\right], \tag{A.10}$$

$$[\underline{\delta}_2(\theta_B), \bar{\delta}_2(\theta_B)] = \left[\frac{1+\beta(\beta(3-2\phi)-4)}{(\beta-1)^2}, 1\right]. \tag{A.11}$$

The challenger's ratification thresholds always exist and are given by:

$$[\underline{\delta}^{j=G}, \bar{\delta}^{j=G}] = \left[\max\left\{0, \frac{1-\beta\,[3+\phi+\beta(\phi-2)(1+\phi)]-\sqrt{M_{j=G}}}{(1+\beta(\phi-1))^2}\right\}, \frac{1-\beta\,[3+\phi+\beta(\phi-2)(1+\phi)]+\sqrt{M_{j=G}}}{(1+\beta(\phi-1))^2}\right\}\right], \tag{A.12}$$

with $M_{j=G} = \beta^2(1-\beta)(1+\phi)\,(1+5\phi+\beta\,[\phi(4\phi-5)-1])$.

Proof of Proposition A.3
(i) Ratification interval for $i = B$:
Threshold values follow from (9), which is a concave function and has two roots. The upper ratification threshold is equal to 1 because it corresponds to the incumbent's non-cooperative emission choice and thus makes them equally well off as without a treaty.

The lower ratification threshold is always positive, within the assumed parameter ranges:

$$\underline{\delta}^{i=B} = \frac{1+\beta(\phi-1)[4+\beta(\phi-3)]}{(1+\beta(\phi-1))^2} > 0$$
$$1 > \beta(1-\phi)[4+\beta(\phi-3)],$$

where the right hand side at its largest takes the value 0.555, and therefore is always smaller than 1.

(ii) Ratification interval for country 2:
Threshold values follow from (13), which is a concave function and has two roots. The upper ratification threshold is equal to 1 because it corresponds to country 2's non-cooperative emission choice and thus makes them equally well off as without a treaty.

The lower ratification threshold is always positive:

$$\underline{\delta}_2 = \frac{1+\beta[2\beta(1-\phi)+\beta-4]}{(\beta-1)^2} > 0$$
$$\underbrace{1+3\beta^2}_{\in[1,1.07]} > \underbrace{4\beta+2\beta^2\phi}_{\in[0,0.64]}$$

Given that the two intervals never overlap, this is always true.

(iii) Ratification interval for $j = G$:
Threshold values follow from (9), which is a concave function and has two roots. They exist if $M_{j=G}$ is non-negative:

$$M_{j=G} = \underbrace{\beta(1+\phi)(1-\beta)}_{>0}\underbrace{\left[1+5\phi+\beta(4\phi^2-5\phi-1)\right]}_{\mathcal{A}}$$

where

$$\mathcal{A} = \underbrace{1-\beta}_{>0}+5\phi\underbrace{(1-\beta)}_{>0}+4\beta\phi^4 > 0$$

and thus $M_{j=G} > 0$, hence the threshold values always exist.

To show when $\underline{\delta}^{j=G}$ is non-negative, consider the value of $\phi \in [0,1]$ that renders $\underline{\delta}^{j=G} = 0$:

$$\underline{\delta}^{j=G}(\phi_0^{j=G}) = 0$$
$$\Rightarrow \phi_0^{j=G} = \frac{2-2\beta-\sqrt{3-4\beta+\beta^2}}{\beta}$$

Knowing that the lower ratification threshold for the green challenger decreases in ϕ (see Proposition A.4), we can show that $\phi_0^{j=G}$ is only restrictive if marginal damages are sufficiently high:

$$\phi_0^{j=G} \leq 1 \;\Rightarrow\; \beta \leq 0.1465$$

Note that this is analogous to the condition in Proposition A.1.

Also, the challenger's upper threshold is never above 1:

$$\bar{\delta}^{j=G} = \frac{1-\beta\left[3+\phi+\beta(\phi-2)(1+\phi)\right]+\sqrt{M_{j=G}}}{\left(1+\beta(\phi-1)\right)^2} \leq 1$$

Rewriting the condition:

$$\frac{\mathcal{B}}{\mathcal{C}} \leq 1 \;\Leftrightarrow\; \mathcal{B}-\mathcal{C} \leq 0$$
$$-4\beta^2\phi^2\left(1+\beta(\phi-1)\right)^2 \leq 0$$

which is always true. □

Proposition A.4 (Stage 3: Comparative Statics with $i = B$)
The following conditions hold for the equilibrium ratification intervals under the condition that they are within the interval $[0, 1]$*:*

$$\frac{d\underline{\delta}^{i=B}}{d\phi} > 0, \qquad \frac{d\bar{\delta}^{i=B}}{d\phi} = 0, \qquad \frac{d\underline{\delta}^{j=G}}{d\phi} < 0, \qquad \frac{d\bar{\delta}^{j=G}}{d\phi} < 0.$$

Proof of Proposition A.4
(i) For the brown incumbent's thresholds:

$$\frac{d\underline{\delta}^{i=B}}{d\phi} = \frac{2\beta\overbrace{(\beta-1)}^{<0}\overbrace{(\beta(\phi-1)-1)}^{<0}}{\underbrace{\left(1+\beta(\phi-1)\right)^3}_{>0}} > 0 \tag{A.13}$$

The upper ratification threshold of the brown incumbent is not a function of ϕ.

(ii) For the green challenger's thresholds:

- $\frac{d\underline{\delta}^{j=G}}{d\phi} < 0$

$$\frac{d\underline{\delta}^{j=G}}{d\phi} = \frac{\beta(\beta-1)\Big[3\sqrt{M_{j=G}} + \beta(3+\beta(\phi^2-15\phi-8)+}{(1+\beta(\phi-1))^3\sqrt{M_{j=G}}}$$
$$\frac{\beta^2(5+10\phi-5\phi^2-2\phi^3)-5\sqrt{M_{j=G}}+\phi(5+\sqrt{M_{j=G}})\Big]}{(1+\beta(\phi-1))^3\sqrt{M_{j=G}}} \tag{A.14}$$

Note that the denominator is positive, as seen in (A.13). In the numerator, given that $\beta(\beta-1) < 0$, I will show that the term in square brackets is positive. Rewriting this term:

$$\sqrt{M_{j=G}}\overbrace{(3-5\beta+\beta\phi)}^{\mathcal{A}} + \beta\overbrace{\left(3-8\beta+5\beta^2\right)}^{\mathcal{B}} + \phi\overbrace{\left(5\beta-15\beta^2+10\beta^3\right)}^{\mathcal{C}}$$
$$+\phi^2\underbrace{\left(\beta^2-5\beta^3\right)}_{\mathcal{D}} + \phi^3\underbrace{\left(-2\beta^3\right)}_{\mathcal{E}}$$

where

$$\begin{aligned}
&\mathcal{A} = 3-5\beta+\beta\phi > 0 \quad \text{for } \beta < \frac{3}{5-\phi} \in [0.6, 0.71],\ \text{depending on } \phi \in [0, \bar{\phi}^B].\\
&\mathcal{B} = 3-8\beta+5\beta^2 > 0 \quad \text{for } \beta < 0.6\\
&\mathcal{C} = 5\beta-15\beta^2+10\beta^3 > 0 \quad \text{for } \beta < 0.5\\
&\mathcal{D} = \beta^2-5\beta^3 > 0 \quad \text{for } \beta < 0.2\\
&\mathcal{E} = -2\beta^3 < 0
\end{aligned}$$

All terms but $\mathcal{E}$ are positive. However, the negative impact of $\mathcal{E}$ is covered, e.g. by term $\mathcal{C}$ as follows:

$$\phi\mathcal{C} - \phi^3\mathcal{E} = \phi(5\beta-15\beta^2+10\beta^3-(2\beta^3\phi^2)) = \phi(\overbrace{5\beta-15\beta^2}^{>0 \text{ for } \beta<\frac{1}{3}} + \beta^3\overbrace{(10-2\phi^2)}^{>0}) > 0$$

Consequently, the term in square brackets is positive, making the numerator of (A.14) negative and hence the whole expression negative.

- $\frac{d\bar{\delta}^{j=G}}{d\phi} < 0$

$$\frac{d\bar{\delta}^{j=G}}{d\phi} = \frac{\beta(\beta-1)\Big[3\sqrt{M_{j=G}} + \beta(-3+\beta(8+15\phi-\phi^2)+}{(1+\beta(\phi-1))^3\sqrt{M_{j=G}}}$$
$$\frac{\beta^2(-5-10\phi+5\phi^2+2\phi^3)-5\sqrt{M_{j=G}}+\phi(-5+\sqrt{M_{j=G}})\Big]}{(1+\beta(\phi-1))^3\sqrt{M_{j=G}}} \tag{A.15}$$

The denominator is positive, as seen in (A.13). Again, given that $\beta(\beta-1)<0$, I will show that the term in square brackets is positive. Rewriting this term:

$$\begin{aligned}
\mathcal{A} + [3-5\beta+\beta\phi]\sqrt{M_{j=G}} &> 0 \\
\sqrt{M_{j=G}} &> -\frac{\mathcal{A}}{[3-5\beta+\beta\phi]} \\
M_{j=G} - \left(\frac{\mathcal{A}}{[3-5\beta+\beta\phi]}\right)^2 &> 0
\end{aligned}$$

which simplifies to

$$\frac{1}{(\cdot)^2}\left(4\beta^2\phi\underbrace{(1-\beta(\phi-1))^3}_{>0}\underbrace{[6+5\phi-\overbrace{\beta(10+9\phi+\phi^2)}^{\max=3}]}_{>0}\right) > 0.$$

Therefore, the term in square brackets is negative, making the whole expression negative.

□

Lemma A.2 (Ordering of Countries' Ratification Intervals with $i = B$)
There exists a threshold value $\bar{\phi}^B$ for ratification intervals to touch, i.e. at which $\underline{\delta}_i = \bar{\delta}_j$. Then, if:

$$\phi \leq \bar{\phi}^B(\beta) \;\in [0.768, 0.8) \quad \textit{for } \beta \in [0, 0.15], \tag{A.16}$$

a common ratification interval exists.

Proof of Lemma A.2
The two scenarios are separated at the point where the ratification intervals touch, that is,

where $\underline{\delta}_i = \bar{\delta}_j$. Solving this for ϕ yields:

$$\bar{\phi}^B = \frac{\sqrt[3]{3\sqrt{3}\sqrt{(\beta-1)^3\beta^6(\beta(\beta(7\beta-15)+41)-25)} - 2(\beta-1)^2\beta^3(4\beta-13)-}}{3\beta^2}$$
$$\frac{\frac{(\beta-1)\beta^2(5\beta+1)}{\sqrt[3]{3\sqrt{3}\sqrt{(\beta-1)^3\beta^6(\beta(\beta(7\beta-15)+41)-25)}-2(\beta-1)^2\beta^3(4\beta-13)}} + 4(\beta-1)\beta}{3\beta^2} \tag{A.17}$$

where $\frac{d\bar{\phi}^B}{d\beta} < 0$, i.e., the higher environmental damages, the smaller the range for common ratification. □

Lemma A.3 (Restrictions on Shock Density)
For reelection probabilities to be interior in $(0, 1)$*, the variance in the popularity shock is restricted to:*

$$\sigma < \bar{\sigma} = \min\left\{\frac{1-z}{\Delta W_M^l}, \frac{z}{|\Delta W_M^l|}\right\}, \tag{A.18}$$

which is most restrictive for the case $l \in \{A, B, C, D\}$ *for which* $|\Delta W_M^l|$ *is highest.*

Proof of Lemma A.3
The reelection probability has to be interior, that is $p^l \in (0, 1)$. Note that the reelection probability is increased versus the incumbency advantage if $\Delta W_M^l > 0$ and vice versa for $\Delta W_M^l < 0$. In the first case, we thus have to ensure that:

$$\sigma \Delta W_M^l + z < 1$$
$$\sigma < \frac{1-z}{\Delta W_M^l},$$

while for $\Delta W_M^l < 0$ it has to hold that:

$$\sigma \Delta W_M^l + z > 0$$
$$\sigma < \frac{z}{-\Delta W_M^l} = \frac{z}{|\Delta W_M^l|}.$$

Note that for high values of ΔW_M^l these conditions become harder to fulfil (since the upper limit is lower). Therefore, whichever case A–D leads to the highest value of difference in median voter welfare $|\Delta W_M^l|$ will be restrictive for the shock density and is thus defined as $\bar{\sigma}$.

□

Lemma A.4 (Polarisation Thresholds with $i = G$)
In addition to $\bar{\phi}^G$, which governs whether or not a common ratification interval exists as defined in Lemma A.1, the threshold value $\bar{\phi}^G_{FB}$ determines whether the first-best outcome, i.e. the no-election treaty parameter, leads to a consensus treaty.

It holds that $\bar{\phi}^G_{FB} < \bar{\phi}^G$ in the relevant parameter range for β.

Proof of Lemma A.4
Firstly, the parameter $\bar{\phi}^G_{FB}$ is defined as the polarisation level at which the no-election treaty, as defined by (5), is just within the area C, that is when $\hat{\delta}_i = \underline{\delta}_j$ holds. For a green incumbent, (5) is given by:

$$\hat{\delta}_i(1+\phi) = \frac{1+\beta(1+\phi)[\beta(2+\phi)-3]}{[1-\beta(1+\phi)]^2}, \tag{A.19}$$

where

$$\frac{d\hat{\delta}_i(1+\phi)}{d\phi} = \frac{\overbrace{\beta[1+\beta\phi-\beta^2(1+\phi)]}^{>0}}{\underbrace{[\beta(1+\phi)-1]^3}_{<0}} < 0.$$

Therefore, as polarisation increases, (A.19) decreases and thus, $\bar{\phi}^G_{FB}$ is the highest level of polarisation which allows for a first-best treaty.

The threshold parameter $\bar{\phi}^G$ is defined in (A.4). For the parameter range $\beta \in (0, 0.15]$ the two threshold parameters take the following values:

$$\bar{\phi}^G_{FB} \in [0.133, 0.136] < \bar{\phi}^G \in [0.2, 0.206].$$

□

Lemma A.5 (Polarisation Thresholds with $i = B$)
In addition to $\bar{\phi}^B$, which governs whether or not a common ratification interval exists as defined in Lemma A.2, the threshold value $\bar{\phi}^B_{FB}$ determines whether the first-best outcome, i.e. the no-election treaty parameter, leads to a consensus treaty.

It holds that $\bar{\phi}^B_{FB} < \bar{\phi}^B$ in the relevant parameter range for β.

Proof of Lemma A.5
Firstly, the parameter $\bar{\phi}^B_{FB}$ is defined as the polarisation level at which the no-election treaty, as defined by (5), is just within the area C, that is when $\hat{\delta}_i = \bar{\delta}_j$ holds. For a brown

incumbent, (5) is given by:

$$\hat{\delta}_i(1-\phi) = \frac{1-\beta(1-\phi)[3-\beta(2-\phi)]}{[1-\beta(1-\phi)]^2}, \tag{A.20}$$

where

$$\frac{d\hat{\delta}_i(1-\phi)}{d\phi} = \frac{\beta\overbrace{[1-\beta\phi-\beta^2(1-\phi)]}^{>0}}{\underbrace{[1-\beta(1-\phi)]^3}_{>0}} > 0.$$

Thus, as polarisation increases, (A.20) increases and thus, $\bar{\phi}^B_{FB}$ is the highest level of polarisation which allows for a first-best treaty.

The threshold parameter $\bar{\phi}^B$ is defined in (A.17). For the parameter range $\beta \in (0, 0.15]$ the two threshold parameters take the following values:

$$\bar{\phi}^B_{FB} \in [0.645, 0.679] < \bar{\phi}^B \in [0.768, 0.798].$$

□

B Proofs

Proof of Proposition 1

1. To show that (15a) is true, note that for $\phi = 0$ it holds that $\underline{\delta}^{i=G} = \underline{\delta}^{j=B}$ (because in this case $\theta_i = \theta_j$) and thus $\Delta\underline{\delta}^{i=G} = 0$. As stated in Proposition A.2, the incumbent's lower threshold decreases in ϕ as shown in (A.6), while the challenger's lower threshold increases in ϕ as shown in (A.7). It thus follows that $\Delta\underline{\delta}^{i=G} \leq 0$.

   By the same reasoning, (15b) is true. At $\phi = 0$, it holds that $\bar{\delta}^{i=G} = \bar{\delta}^{j=B}$. Then, $\bar{\delta}^{i=G}$ is constant in ϕ, while $\bar{\delta}^{j=B}$ decreases with increasing ϕ, as shown in (A.9), meaning that $\Delta\bar{\delta}^{i=G} \geq 0$.

2. To show that (16a) holds true, first note that if $\phi = 0$, $\underline{\delta}^{i=G} = \underline{\delta}_2(\theta_G)$ because the two share the same environmental preference parameter. Then, as ϕ increases, the two values diverge as follows:

$$\frac{d\underline{\delta}_2}{d\phi} = \frac{2\beta^2}{(\beta-1)^2} \geq 0$$
$$\frac{d\underline{\delta}^{i=G}}{d\phi} \leq 0 \qquad \text{as shown in (A.6).}$$

   Therefore, for any $\phi \in [0,1]$, $\underline{\delta}^{i=G} \leq \underline{\delta}_2(\theta_G)$.

   To show that (16b) holds true, a couple of steps are necessary. First, comparing the denominators of the two elements:

$$\bar{\delta}_2 = \frac{\mathcal{A}}{(\beta-1)^2}$$
$$\bar{\delta}^{j=B} = \frac{\mathcal{B}}{(\beta-1+\beta\phi)^2} \ .$$

   Since $(\beta-1)^2 \geq (\beta-1+\beta\phi)^2$, it holds that $\frac{1}{(\beta-1)^2} \leq \frac{1}{(\beta-1+\beta\phi)^2}$. Consequently, if $\mathcal{A} \leq \mathcal{B}$ is true, then $\bar{\delta}_2 \leq \bar{\delta}^{j=B}$ follows and (16b) holds true. I therefore now show that $\mathcal{A} \leq \mathcal{B}$:

$$\mathcal{B} - \mathcal{A} \geq 0$$
$$\beta(\phi+1) - \beta^2\left(\phi^2 + 3\phi + 1\right) - \sqrt{M_{j=B}} \geq 0$$
$$\phi\underbrace{(8-12\beta)}_{>6.2} + \phi^2\underbrace{(2\beta - 4 - 12\beta^2)}_{>-4} + \phi^3\underbrace{(4\beta - 10\beta^2)}_{>0} + \phi^4\underbrace{(-\beta^2)}_{>-0.0225} + 2\beta\underbrace{(1-\beta)}_{>0.85} \geq 0 \ .$$

In the last line, I evaluate all expressions at the value which renders the condition most strict. Note that the positive terms in brackets are in any case sufficient to cover the negative terms in brackets, even more so when taking into account the multiplications with ϕ.

$\square$

Proof of Proposition 2

1. To show that (17a) is true, note that for $\phi = 0$ it holds that $\underline{\delta}^{i=B} = \underline{\delta}^{j=G}$ and thus $\Delta\underline{\delta}^{i=B} = 0$. As stated in Proposition A.4, the incumbent's lower threshold increases in ϕ as shown in (A.13), while the challenger's lower threshold decreases in ϕ as shown in (A.14). It thus follows that $\Delta\underline{\delta}^{i=B} \geq 0$.

   By the same reasoning, the second relation of this Proposition, that is, (17b), is true. At $\phi = 0$, it holds that $\bar{\delta}^{i=B} = \bar{\delta}^{j=G}$. Then, $\bar{\delta}^{i=B}$ is constant in ϕ, while $\bar{\delta}^{j=G}$ decreases with increasing ϕ, as shown in (A.15), meaning that $\Delta\bar{\delta}^{i=B} \geq 0$.

2. Proof of (18a): First, if $\phi = 0$, $\underline{\delta}^{i=B} = \underline{\delta}_2(\theta_B)$ because the two share the same preferences. Then, as ϕ increases, the two values diverge as follows:

$$\frac{d\underline{\delta}_2}{d\phi} = \frac{-2\beta^2}{(\beta-1)^2} \leq 0$$
$$\frac{d\underline{\delta}^{i=B}}{d\phi} > 0 \qquad \text{as shown in (A.13).}$$

   Therefore, for any $\phi \in [0, \bar{\phi}^B]$, $\underline{\delta}^{i=B} \geq \underline{\delta}_2(\theta_B)$ and thus (18a) holds.

   To show that (18b) holds true, a couple of steps are necessary. First, comparing the denominators of the two elements:

$$\bar{\delta}_2 = \frac{\mathcal{A}}{(\beta-1)^2}$$
$$\bar{\delta}^{j=G} = \frac{\mathcal{B}}{(\beta-1-\beta\phi)^2} \; .$$

   Since $(\beta-1)^2 \geq (\beta-1-\beta\phi)^2$, reformulating $\underline{\delta}^2 \geq \delta^{j=G}$ yields:

$$\frac{\mathcal{A}}{(\beta-1)^2} \geq \frac{\mathcal{B}}{(\beta-1-\beta\phi)^2}$$
$$\frac{\mathcal{A}}{\mathcal{B}} \geq \frac{(\beta-1)^2}{\beta-1-\beta\phi)^2} \geq 1$$
$$\Rightarrow \; \mathcal{A} \geq \mathcal{B}$$

We therefore now show that $\mathcal{A} \geq \mathcal{B}$:

$$\begin{aligned}\mathcal{A} - \mathcal{B} &\geq 0\\ \beta(1-\phi) + \beta^2\left(3\phi + 3 - \phi^2 - 1\right) + \sqrt{M_{j=G}} &\geq 0\\ \underbrace{(8-20\beta+12\beta^2)}_{>0} + \phi\underbrace{(4+2\beta-10\beta^2)}_{>0} + \phi^2\underbrace{(2\beta+2\beta^2)}_{>0} + \phi^3\underbrace{(-\beta^2)}_{<0} &> 0.\end{aligned}$$

Note that $2\beta^2\phi^2 > \beta^2\phi^3$ and thus the whole expression is strictly positive.

□

Proof of Proposition 3

Following (19), we know that the incumbent is reelected whenever $\Delta W_M \geq \Omega$. Denoting $\Delta W_M = X$, we are interested in $P(X \geq \Omega) = P(\Omega \leq X) = F_\Omega(X)$. Given the distribution of Ω, we know that $F(X) = \sigma X + z$. It thus follows that $F(0) = z$ (in the absence of policy differences, the reelection probability is equal to the incumbency advantage) and generally $F(\Delta W_M) = \sigma \Delta W_M + z$.

The welfare difference for the median voter (depending on the cases) is then specifically given by:

$$\begin{aligned}\Delta W_M^A &= \delta_i - 0.5(1+\delta_i^2) + \beta\left[2 + \delta_i^2\theta_i - \delta_i(2+\theta_i)\right] +\\ &\quad \beta^2\left[\delta_i(1+\theta_i - 0.5\delta_i\theta_i^2) + \theta_j(0.5\theta_j - 1) - 1\right]\\ \Delta W_M^B &= 0.5(1-\delta_i)^2 + \beta\left[\delta_i(2+\theta_i-\delta_i+\beta^2\left[1+\theta_i+0.5\theta_i^2(\delta_i^2-1) - \delta_i(1+\theta_i)\right]\theta_i) - 2\right]\\ \Delta W_M^C &= 0\\ \Delta W_M^D &= \beta^2\left[\theta_i(1-0.5\theta_i) + \theta_j(0.5\theta_j - 1)\right].\end{aligned}$$

□

Proof of Proposition 4

(i) 1. To prove that the first point is true, we will show that $W_i^A(\delta = \underline{\delta}_j) > W_i^A(\delta = \bar{\delta}_j)$, meaning that there exists a point in the lower area A that yields a higher expected welfare than the highest point in the upper area A. Rewriting:

$$W_i^A(\underline{\delta}_j) - W_i^A(\bar{\delta}_j) = p^A\left[B(\tilde{e}_1(\underline{\delta}_j)) - \theta_i D(\tilde{E}(\underline{\delta}_j)) + R\right] + (1-p^A)\left[\hat{W}_i(\theta_j)\right] - \left(p^A\left[B(\tilde{e}_1(\bar{\delta}_j)) - \theta_i D(\tilde{E}(\bar{\delta}_j)) + R\right] + (1-p^A)\left[\hat{W}_i(\theta_j)\right]\right)$$

$$= p^A \underbrace{\left[B(\tilde{e}_1(\underline{\delta}_j)) - B(\tilde{e}_1(\bar{\delta}_j)) + \theta_i D(\tilde{E}(\bar{\delta}_j)) - \theta_i D(\tilde{E}(\underline{\delta}_j))\right]}_{\mathcal{A}},$$

where

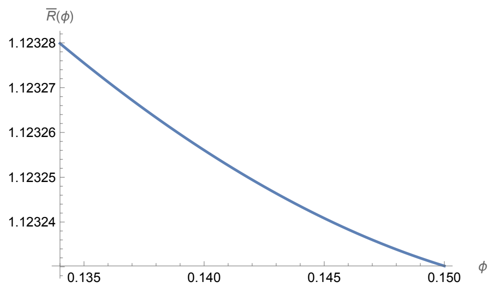
$$\mathcal{A} = \underbrace{\left(8\beta\phi - 4\beta^2\phi(2+\phi)\right)}_{\alpha}\ \underbrace{\sqrt{M_{j=B}}}_{\geq 0 \text{ for } \phi \leq \bar{\phi}^G},$$

and $\alpha > 0$ if $2 > \beta(2+\phi)$, which is strictest at $\beta = 0.15$ and $\phi = \bar{\phi}^G$, where it holds, making the whole expression positive.

2. The office rent which divides consensus and differentiation treaties is defined by (26) for different ranges of the polarisation spectrum. As an illustrative example, consider the second line of (26):

$$\bar{R}^G(\phi) = \arg\min |W^C(\underline{\delta}_j) - W^A(\underline{\delta}_j)| \qquad \text{for } \bar{\phi}^G_{FB} < \phi < \bar{\phi}^G_A$$
$$\Rightarrow \quad W^C(\underline{\delta}_j) = W^A(\underline{\delta}_j)$$

The office rent which solves the above equality is a function of polarisation and for the numerical example can be plotted as follows:



Computing this for all three polarisation ranges then generates the $\bar{R}^G$ function as depicted in Figure 5a.

3. Whether the consensus treaty is of type *first-best* or *compromise* simply depends on availability. By definition, a first-best treaty is the global maximum of the

W^C function and therefore will be picked if available. Availability is determined by whether $\hat{\delta}_i$ is contained within area C, which is true for:

$$\hat{\delta}_i \geq \underline{\delta}_j \;\Rightarrow\; \hat{\delta}_i - \underline{\delta}_j \geq 0$$

$$\frac{2\beta^2\phi(2+\phi) - 4\beta\phi + \sqrt{\beta^2(\beta-1)(\phi-1)(1-5\phi+\beta(4\phi^2+5\phi-1))}}{(\beta(1+\phi)-1)^2} \geq 0$$

The sign of this expression is determined by the numerator. The threshold value of polarisation for which $\hat{\delta}_i = \underline{\delta}_j$ holds is denoted by $\bar{\phi}^G_{FB}$. An explicit expression for $\bar{\phi}^G_{FB}(\beta)$ is not possible, but numerical solutions for the range $\beta \in (0, 0.15]$ are $\bar{\phi}^G_{FB} \in (0.1316, 0.1357)$, where $\frac{\bar{\phi}^G_{FB}}{d\beta} < 0$ in this range. Therefore, for $\phi \geq \bar{\phi}^G_{FB}$, the consensus treaty is first-best, and a compromise treaty otherwise.

Note that for a compromise treaty it holds that $\hat{\delta}_i < \underline{\delta}_j$, therefore $\frac{dW^C}{d\delta}|_{\delta=\underline{\delta}_j} < 0$, meaning that the corner solution is the welfare-maximising choice for the incumbent, i.e. $\delta^*_i = \underline{\delta}_j + \epsilon$.

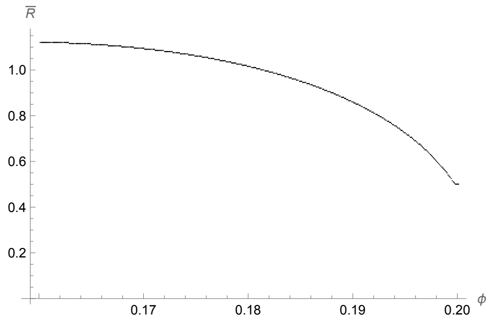
4. In the range in which the incumbent chooses between a compromise and a distinction treaty, i.e., $\phi \in (\bar{\phi}^G_{FB}, \bar{\phi}^G)$, the specific choice of distinction treaty is not trivial. It is possible that the global maximum of the function W^A (i.e., $\delta^*_{i,A}$) lies within or outside of the range of case A. If $\delta^*_{i,A} \leq \underline{\delta}_j$, then $\delta^*_i = \delta^*_{i,A}$, otherwise, $\delta^*_i = \underline{\delta}_j + \epsilon$. However, the value of $\delta^*_{i,A}$ cannot be computed analytically in an explicit fashion. Therefore, we resort to a numerical approach, which is detailed in the following.

   Whether or not the global maximum is available, depends both on polarisation and on the office rent. Crucially, the office rent does not only affect the level of W^A directly, but also indirectly via the reelection probability p^A. Consequently, it is not possible to generally compute a threshold value for ϕ at which $\delta^*_{i,A}$ becomes available, since this depends on the level of the office rent. The following steps are necessary for the numerical computation of $\bar{R}(\phi)$ in the range $\phi \in (\bar{\phi}^G_A, \bar{\phi}^G)$:

   i. $\bar{R}(\phi)$ is the level of office rent at which the incumbent is indifferent between choosing the two treaty parameters $\delta^*_{i,A}$ (distinction) and $\underline{\delta}_j + \epsilon$ (compromise), i.e. where the following holds:

   $$W^A(\delta^*_{i,A}, \bar{R}) = W^C(\underline{\delta}_j + \epsilon, \bar{R})$$

   Computing this for the relevant range for ϕ (in steps of 0.0001) yields:



ii. In a second step, we have to find the range of ϕ for which this is the relevant comparison. We are thus looking for a threshold value $\bar{\phi}_A^G$ beyond which the global maximum is available within area A. We can define:

$$\bar{\phi}_A^G: \quad \frac{dW^A}{d\delta}|_{\delta=\underline{\delta}_j, \phi=\bar{\phi}_A^G} = 0 \tag{B.1}$$

This essentially states that at this value $\bar{\phi}_A^G$, the global maximum $\delta_{i,A}^*$ coincides with $\underline{\delta}_j$, and thus the slope of W^A evaluated at this point is zero. Hence, if the derivative evaluated at $\underline{\delta}_j$ takes on a negative value, this means that $\delta_{i,A}^* < \underline{\delta}_j$, which is true for $\phi > \bar{\phi}_A^G$, and vice versa for a positive derivative.

Note that this threshold value depends on R. Thus, it can only be computed for a specific values of the office rent, however, for $R \in (0, 3)$, $\bar{\phi}_A^G \in (0.1345, 0.1648)$ with $\frac{d\bar{\phi}_A^G}{dR} > 0$.

In combination, for any level of the office rent, these two steps provide the range $(\bar{\phi}_A^G, \bar{\phi}^G)$ in which the global maximum of area A, i.e. $\delta_{i,A}^*$ is available, and then the corresponding level of $\bar{R}$ which in this range separates compromise from distinction treaties.

(ii) While we cannot solve explicitly for the optimal treaty parameter within area A ($\delta_{i,A}^*$), it is implicitly defined by the following expression:

$$\frac{dW_i^A}{d\delta_i} = \frac{dp^A}{d\delta_i}\left[\tilde{W}_i(\delta_i) - \hat{W}_i(\theta_j) + R\right] + \frac{d\tilde{W}_i}{d\delta_i}p^A = 0. \tag{B.2}$$

To see that $\delta_{i,A}^* \in (\hat{\delta}_i, \delta_M^*)$, we will evaluate (B.2) at $\hat{\delta}_i$ and δ_M^* and show that it is increasing in the former and decreasing in the latter, meaning that graphically, $\delta_{i,A}^*$ is located in between the two. Two prerequisites are necessary to show this.

First, note that the treaty parameter within case A that maximises the reelection probability coincides with the median voter's optimal treaty parameter:

$$\delta_A^{max} = \delta_M^* = \frac{1+\beta(\beta(1+\theta_i)-2-\theta_i)}{(1-\beta\theta_i)^2} \quad \text{where} \quad \frac{dp^A}{d\delta_i} = \begin{cases} >0 & \text{for } \delta_i < \delta_A^{max} \\ <0 & \text{for } \delta_i > \delta_A^{max}. \end{cases}$$

Second, it holds that $\hat{\delta}_i < \delta_M^*$, because:

$$\delta_M^* - \hat{\delta}_i = \frac{2\beta\phi - \beta^2\phi(\phi+2)}{(\beta(\phi+1)-1)^2} > 0,$$

because $2 > \beta(2+\phi)$ as seen in the proof of part (i) of this Proposition.

Now, note that $\hat{\delta}_i$ is defined by $\frac{d\tilde{W}_i}{d\delta_i} = 0$. The derivative (B.2) evaluated at $\hat{\delta}_i$ thus becomes:

$$\frac{dW_i^A}{d\delta_i}\Big|_{\delta=\hat{\delta}_i} = \frac{dp^A}{d\delta_i}\left[\tilde{W}_i(\delta_i) - \hat{W}_i(\theta_j) + R\right] > 0. \tag{B.3}$$

This is true because:

$$\tilde{W}_i(\delta_i) - \hat{W}_i(\theta_i) = \frac{\beta^2 \overbrace{\left(0.5-\beta+0.5\beta^2\right)}^{>0}(\phi+1)^2}{(\beta(\phi+1)-1)^2} > 0,$$

and $\hat{W}_i(\theta_i) > \hat{W}_i(\theta_j)$. Also as shown, $\hat{\delta}_i < \delta_M^*$ implies that $\frac{dp^A}{d\delta_i} > 0$. The positive sign of (B.3) implies that $\hat{\delta}_i < \delta_{i,A}^*$.

In a next step, note that δ_M^* is defined by $\frac{dp^A}{d\delta_i} = 0$, as shown above. The derivative (B.2) evaluated at δ_M^* thus becomes:

$$\frac{dW_i^A}{d\delta_i}\Big|_{\delta=\delta_M^*} = \underbrace{\frac{d\tilde{W}_i}{d\delta_i}}_{<0} p^A < 0, \tag{B.4}$$

since $\frac{d\tilde{W}_i}{d\delta_i} = 0$ and $\hat{\delta}_i < \delta_M^*$. The negative sign of (B.4) thus implies that $\delta_M^* > \delta_{i,A}^*$.

□

Proof of Proposition 5

(i) 1. To prove that the first point is true, we will show that $W_i^B(\delta = \underline{\delta}_i) \leq W_i^C(\delta = \underline{\delta}_i)$, meaning that there always exists a point in area C which yields a weakly higher expected welfare for the incumbent than the highest level in area B, thus they never choose a treaty in area B. First, rewriting:

$$\begin{aligned} W_i^C - W_i^B &= B(\tilde{e}_1) - \theta_i D(\tilde{E}) + p^C R - \\ &\qquad \left[p^B (B(\hat{e}_{1,i}) - \theta_i D(\hat{E}) + R) + (1-p^B)(B(\tilde{e}_1) - \theta_i D(\tilde{E}))\right] \\ &= \left[B(\tilde{e}_1) - \theta_i D(\tilde{E})\right] p^B + R(p^C - p^B) - \left[B(\hat{e}_{1,i}) - \theta_i D(\hat{E})\right] p^B \\ &= p^B \Bigg[\underbrace{\tilde{W}_i - \hat{W}_i}_{\mathcal{B}}\Bigg] + R \Bigg[\underbrace{p^C - p^B}_{\mathcal{C}}\Bigg] . \end{aligned}$$

We will now evaluate this difference at $\delta = \underline{\delta}_i$, where cases B and C meet. Now note that by definition of ratification threshold values, $\mathcal{B}$ is equal to zero: at $\underline{\delta}_i$, the incumbent is indifferent between ratifying or not, making the two values exactly equal.

Therefore, to show that $W_i^B(\delta = \underline{\delta}_i) \leq W_i^C(\delta = \underline{\delta}_i)$ holds, $\mathcal{C}$ has to be non-negative:

$$\left.(p^C - p^B)\right|_{\delta = \underline{\delta}_i} = \frac{4 - 4\phi + \beta\left(-2\phi^2 + 10\phi - 8\right) + \beta^2\left(2\phi^2 - 6\phi + 4\right)}{(1 + \beta(\phi - 1))^2} .$$

Given that the denominator is quadratic, the numerator determines the sign of the expression.

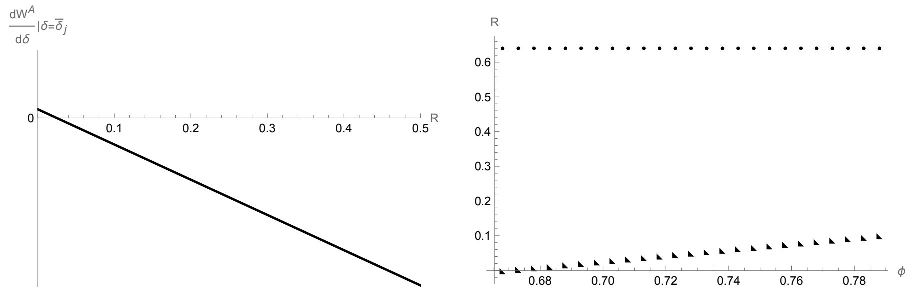
$$\underbrace{4 - 4\phi + \beta\left(-2\phi^2 + 10\phi - 8\right)}_{\mathcal{D}} + \beta^2 \underbrace{\left(2\phi^2 - 6\phi + 4\right)}_{\geq 0 \text{ for } \phi \leq 1}$$

Note that $\mathcal{D}$ decreases as ϕ increases (within the given range) and equals $2\beta - 2\beta^2$ at $\phi = 1$, meaning that $\mathcal{D}$ is non-negative for all values $\beta \in (0, 0.15]$. Consequently, a treaty in area B is always weakly dominated by a treaty in area C.

2. The office rent which divides consensus and differentiation treaties is defined by (27) for different ranges of the polarisation spectrum. Contrary to the case

of a green incumbent, here, we **do not** have to consider the global maximum within area A for comparison. In the following, I show that this is true for the polarisation range $\phi \in (\bar{\phi}^B_{FB}, \bar{\phi}^B)$, where the choice is between a compromise and a differentiation treaty.

Analogously to the proof of Proposition 4 part (i), 4., we can numerically compute the level of the office rent $\bar{R}$ which separates a compromise treaty with $\bar{\delta}_j$ and a differentiation treaty with $\delta^*_{i,A}$, displayed in Figure 15b with the black dots. Now in a second step, we have to check whether the global maximum is even available within area A: computing the derivative of W^A and evaluating at $\delta = \bar{\delta}_j$ indicates whether $\bar{\delta}_j \gtrless \delta^*_{i,A}$, depicted in Figure 15a. Here we see that the global maximum is only available for very small values of the office rent (where the derivative is positive). Numerically solving for the root of this derivative for the relevant ϕ range gives rise to the highest R values for which $\delta^*_{i,A}$ is in area A, plotted as black triangles in Figure 15b.



(a) $\phi = 0.7$, slope of W^A at threshold $\bar{\delta}_j$

(b) Dots: R which separates $W^A(\delta^*_{i,A})$ and $W^C(\bar{\delta}_j)$. Triangles: threshold R at which $\delta^*_{i,A} \geq \bar{\delta}_j$

Figure 15: Global maximum of W^A is not an equilibrium in $\delta \in (\bar{\delta}_j, 1)$

In this combined plot, we thus see that the R values for which $\delta^*_{i,A}$ is in area A are clearly smaller than the R values for which the incumbent would be indifferent between a distinction and compromise treaty. In other words, the R levels which render the incumbent indifferent correspond to $\delta^*_{i,A}$ values which are outside of area A, and thus are not viable choices for a distinction treaty. Hence, in case of a brown incumbent, for distinction treaties it holds that the optimal treaty is given by $\delta^*_i = \bar{\delta}_j$.

3. Analogously to Proposition 4, whether the consensus treaty is of type *first-best* or *compromise* simply depends on availability. By definition, a first-best treaty is the global maximum of the W^C function and therefore will be picked if available.

Availability is determined by whether $\hat{\delta}_i$ is contained within area C, which is true for:

$$\underline{\delta}_i \leq \hat{\delta}_i \leq \bar{\delta}_j$$

First:

$$\hat{\delta}_i - \underline{\delta}_i = \frac{1}{(\cdot)^2}\left[\beta(1-\phi+\beta(\phi-1))\right] > 0 \qquad \text{for } \beta < 1, \text{ which is always true.}$$

Second:

$$\hat{\delta}_i - \bar{\delta}_j = \frac{1}{(\cdot)^2}\left[\mathcal{A} - \sqrt{M_{j=G}}\right], \tag{B.5}$$

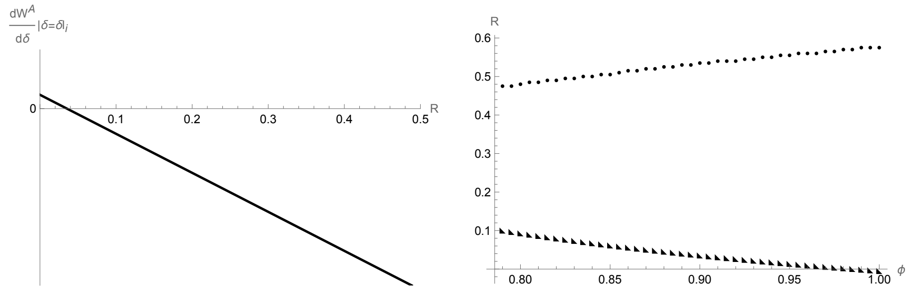
where the term in square brackets is negative whenever:

$$M_{j=G} - \mathcal{A}^2 > 0$$
$$\beta^2(-4\phi^4 + 12\phi^3 - 15\phi^2 + 6\phi + 1) + \beta(-12\phi^3 + 26\phi^2 - 12\phi - 2) - 11\phi^2 + 6\phi + 1 > 0.$$

We cannot solve this explicitly for $\phi(\beta)$, however, numerically for relevant β values. We find the threshold value for the availability of the first-best treaty, i.e. $\bar{\phi}^B_{FB}$, at the point at which the above inequality holds with equality. Thus, for $\beta \in (0, 0.15]$, $\bar{\phi}^B_{FB} \in (0.6448, 0.6793)$ with $\frac{d\phi^B_{FB}}{d\beta} < 0$.

(ii) The office rent which separates insurance and differentiation treaties is defined by (27) for $\phi > \bar{\phi}^B$. Again, contrary to the case of a green incumbent, here, we **do not** have to consider the the global maximum within area A for comparison. In the following, I show that this is true for the polarisation range $\phi \in (\bar{\phi}^B, 1)$, where the choice is between an insurance treaty and a differentiation treaty. The argumentation is perfectly analogous to (i), 2.

Figure 16a shows the slope of W^A evaluated at the threshold value $\underline{\delta}_i$, which in this case is the boundary value between area D and A. Again, for very small levels of R the slope is positive, meaning that $\delta^*_{i,A} > \underline{\delta}_i$. The threshold values for R, for which $\delta^*_{i,A}$ is available in area A are depicted as triangles in Figure 16b. In the same figure, the dots refer to the threshold value $\bar{R}$, resulting from the comparison of an insurance treaty with $\bar{\delta}_j$ (separating area C and area D) and a differentiation treaty with $\delta^*_{i,A}$. Again, we can see that these office rent levels clearly diverge. Thus, for the necessary levels of R at which the incumbent would be indifferent between a compromise and an insurance treaty, the global optimum of W^A is not available as a treaty parameter. Hence, for distinction treaties, it holds that $\delta_i^* = \underline{\delta}_i$.



(a) $\phi = 0.9$, slope of W^A at threshold $\underline{\delta}_i$

(b) Dots: R which separates $W^A(\delta_{i,A}^*)$ and $W^B(\bar{\delta}_j)$. Triangles: threshold R at which $\delta_{i,A}^* \geq \underline{\delta}_i$

Figure 16: Global maximum of W^A is not an equilibrium in $\delta \in (\underline{\delta}_i, 1)$

□

C Extensions

Preference Asymmetry

Here we provide the formal background of the extension with preference asymmetry and how they relate to the main Propositions. More detailed formal proofs of the comparative statics are not provided at this point, but can be graphically confirmed.

Lemma C.1 (Existence of Ratification Interval for Brown Challenger)
The challenger's ratification thresholds exist if:

$$\lambda \leq \bar{\lambda} = \frac{\beta - 1 + 2\mu(1 - \beta - \beta\mu)}{3(\beta - 1) + 2\beta\mu} \tag{C.1}$$

$$\mu \leq \bar{\mu} = \frac{1 - \beta - \sqrt{1 - 4\beta + 3\beta^2}}{2\beta} \tag{C.2}$$

and where it holds that:

$$\frac{d\bar{\lambda}}{d\mu} < 0, \quad \frac{d\bar{\mu}}{d\beta} > 0. \tag{C.3}$$

Proof of Lemma C.1
The challenger's threshold values exist if $M_{j=B}$ is non-negative. Therefore:

$$\underbrace{\beta^2 \overbrace{(\beta - 1)}^{<0} \overbrace{(\lambda - 1)}^{<0}}_{>0} [1 - 2\mu - 3\lambda + \beta(2\mu(1 + \mu) - 1 + \lambda(3 + 2\mu))] \geq 0$$

It thus suffices to consider the term in square brackets to determine the sign:

$$\lambda\,(3(\beta - 1) + 2\beta\mu) \geq \beta - 1 + 2\mu\,(1 - \beta - \beta\mu)$$
$$\lambda \leq \frac{\beta - 1 + 2\mu(1 - \beta - \beta\mu)}{3(\beta - 1) + 2\beta\mu} \equiv \bar{\lambda}$$

Note that the sign switches between the two lines because the denominator is negative for the parameter range considered. This can be shown by assuming the parameters that maximise this expression, i.e. $\beta = 0.15$ and $\mu = 1$ and then $3(\beta - 1) + 2\beta\mu = -2.25 < 0$.

Note that at $\bar{\lambda}(\bar{\mu}) = 0$, meaning that $\mu \leq \bar{\mu}$ ensures that $\bar{\lambda}$ is non-negative.

□

Lemma C.2 (Stage 3: Ratification Intervals with $i = G$)
In the case of a green incumbent, the incumbent's and country 2's ratification thresholds are given as follows:

$$[\underline{\delta}^{i=G}, \bar{\delta}^{i=G}] = \left[\max\left\{0, \frac{1+\beta(1+\mu)\left[\beta(3+\mu)-4\right]}{(\beta+\beta\mu-1)^2}\right\},\ 1\right] \tag{C.4}$$

$$[\underline{\delta}_2(\theta_G), \bar{\delta}_2(\theta_G)] = \left[\frac{1+\beta[2\beta(1+\mu)+\beta-4]}{(\beta-1)^2},\ 1\right] \tag{C.5}$$

The challenger's ratification thresholds exist when $\lambda \leq \bar{\lambda}$ and $\mu \leq \bar{\mu}$. In that case, they are given by:

$$[\underline{\delta}^{j=B}, \bar{\delta}^{j=B}] = \Bigg[\frac{1-\beta\left[3-2\lambda+\mu+\beta(\lambda-1)(2+\mu)\right]-\sqrt{M_{j=B}}}{(\beta+\beta\mu-1)^2}, \min\left\{\frac{1-\beta\left[3-2\lambda+\mu+\beta(\lambda-1)(2+\mu)\right]+\sqrt{M_{j=B}}}{(\beta+\beta\mu-1)^2}, 1\right\}\Bigg] \tag{C.6}$$

where $M_{j=B} = \beta^2(\beta-1)(\lambda-1)\left[1-2\mu-3\lambda+\beta(2\mu(1+\mu)-1+\lambda(3+2\mu))\right]$.

Lemma C.3 (Stage 3: Comparative Statics with $i = G$)
The following conditions hold for the equilibrium ratification intervals under the condition that thresholds exist and that they are within the interval $[0, 1]$*:*

$$\frac{d\underline{\delta}^{i=G}}{d\mu} < 0, \qquad \frac{d\bar{\delta}^{i=G}}{d\mu} = 0, \qquad \frac{d\underline{\delta}^{i=G}}{d\lambda} = 0, \qquad \frac{d\bar{\delta}^{i=G}}{d\lambda} = 0$$

$$\frac{d\underline{\delta}^{j=B}}{d\mu} > 0, \qquad \frac{d\bar{\delta}^{j=B}}{d\mu} < 0, \qquad \frac{d\underline{\delta}^{j=B}}{d\lambda} > 0, \qquad \frac{d\bar{\delta}^{j=B}}{d\lambda} < 0$$

Analogously to the case of symmetry, Proposition 1 holds and as a consequence, cases as illustrated in Figure 3 follow. However, now the two scenarios ares distinguished by whether μ and λ are below threshold values as defined in Lemma C.1.

Lemma C.4 (Stage 3: Ratification Intervals with $i = B$)
In the case of a brown incumbent, ratification thresholds are given as follows:

$$[\underline{\delta}^{i=B}, \bar{\delta}^{i=B}] = \left[\frac{1+\beta(\lambda-1)\left[4+\beta(\lambda-3)\right]}{\left(1+\beta(\lambda-1)\right)^2}, 1\right] \tag{C.7}$$

$$[\underline{\delta}_2(\theta_B), \bar{\delta}_2(\theta_B)] = \left[\frac{1+\beta(2\beta(1-\lambda)+\beta-4)}{(\beta-1)^2}, 1\right] \tag{C.8}$$

The challenger's ratification thresholds always exist and are given by:

$$[\underline{\delta}^{j=G}, \bar{\delta}^{j=G}] = \left[\max\left\{0, \frac{1+\beta\left[\lambda-3-2\mu-\beta(2-\lambda)(1+\mu)\right]-\sqrt{M_{j=G}}}{\left((1+\beta(\lambda-1)\right)^2}\right\},\right. \tag{C.9}$$

$$\left.\frac{1+\beta\left[\lambda-3-2\mu-\beta(2-\lambda)(1+\mu)\right]+\sqrt{M_{j=G}}}{\left((1+\beta(\lambda-1)\right)^2}\right] \tag{C.10}$$

with $M_{j=G} = \beta^2(1-\beta)(1+\mu)\left(1+3\mu+2\lambda+\beta\left[2\lambda(\mu+\lambda-1)-3\mu-1\right]\right)$.

Lemma C.5 (Stage 3: Comparative Statics with $i = B$)
The following conditions hold for the equilibrium ratification intervals under the condition that thresholds exist and that they are within the interval $[0, 1]$*:*

$$\frac{d\underline{\delta}^{i=B}}{d\lambda} > 0, \qquad \frac{d\bar{\delta}^{i=B}}{d\lambda} = 0, \qquad \frac{d\underline{\delta}^{i=B}}{d\mu} = 0, \qquad \frac{d\bar{\delta}^{i=B}}{d\mu} = 0$$

$$\frac{d\underline{\delta}^{j=G}}{d\lambda} \lesseqgtr 0, \qquad \frac{d\bar{\delta}^{j=G}}{d\lambda} < 0, \qquad \frac{d\underline{\delta}^{j=G}}{d\mu} < 0, \qquad \frac{d\bar{\delta}^{j=G}}{d\mu} < 0.$$

Lemma C.6 (Ordering of Countries' Ratification Intervals with $i = B$)
The two scenarios are separated at the point where the ratification intervals touch, i.e. where $\underline{\delta}_i = \bar{\delta}_j$*. Solving this for* μ *yields the following threshold:*

$$\tilde{\mu} = \frac{8-9\lambda-\beta(\lambda-1)(6\lambda-16+\beta\left[8+\lambda(\lambda-5)\right])}{(1+\beta(\lambda-1))^2}, \tag{C.11}$$

where $\frac{d\tilde{\mu}}{d\lambda} < 0$*. The threshold value is non-negative as long as* $\lambda < \lambda_0$*, where* $\lambda_0(\beta)$ *is most restrictive at* $\beta = 0.15$ *and takes a value of* $\lambda_0(0.15) \approx 0.85$*.*

Analogous to symmetry, the findings of Proposition 2 hold and consequently, cases as illustrated by Figure 4 follow.

More sophisticated country 2

Lemma C.7 (Equilibrium Consensus Treaties with Sophisticated Country 2) *For sufficiently low levels of polarisation and a brown incumbent, it holds that:*

$$\bar{\delta}_2^{new} \geq \delta_i^*, \tag{C.12}$$

which implies that any consensus treaty which resulted as an equilibrium in the basic model is still valid with a more sophisticated country 2. The highest polarisation level which allows for this is given by the $\bar{\phi}_{max}^B$*, for which* (C.12) *holds with equality.*

Proof of Lemma C.7

First, the new upper ratification threshold for country 2 is implicitly defined as follows, following (31c):

$$\Delta\mathbb{E}[W_2^C] = 0 \qquad \Rightarrow \left[\underline{\delta}_2^{new}, \bar{\delta}_2^{new}\right]$$

If there is no polarisation, i.e. for $\phi = 0$:

$$\delta_i^* < 1,\ \bar{\delta}_2^{new} = 1$$

Now with increasing polarisation, the optimal treaty value for the brown incumbent increases, whereas the upper participation threshold decreases:

$$\frac{d\delta_i^*}{d\phi} > 0,\ \frac{d\bar{\delta}_2^{new}}{d\phi} < 0$$

Therefore, if $\bar{\delta}_2^{new} - \delta_i^* \geq 0$ at $\phi = \bar{\phi}_{max}^B$, it is true for all other values $\phi < \bar{\phi}_{max}^B$. We find this upper limit by numerically solving (C.12) for $\beta \in (0, 0.15]$ and find that $\bar{\phi}_{max}^B \in (0.4452, 0.4456)$ with $\frac{d\bar{\phi}_{max}^B}{d\beta} < 0$. □

References

Aldy, J. E., S. Barrett, and R. N. Stavins (2003). Thirteen plus one: A comparison of global climate policy architectures. *Climate Policy 3*(4), 373–397.

Alesina, A. and G. Tabellini (1990). A positive theory of fiscal deficits and government debt. *The Review of Economic Studies 57*(3), 403–414.

Andreottola, G. and C. Li (2024). Polarization and policy design. *Journal of Political Economy Microeconomics 2*(1), 1–42.

Austen-Smith, D., W. Dziuda, B. Harstad, and A. Loeper (2019). Gridlock and inefficient policy instruments. *Theoretical Economics 14*(4), 1483–1534.

Baker, S. R., A. Baksy, N. Bloom, S. J. Davis, and J. A. Rodden (2020). Elections, political polarization, and economic uncertainty. Technical report, National Bureau of Economic Research.

Barrett, S. (1994). Self-enforcing international environmental agreements. *Oxford Economic Papers*, 878–894.

Barrett, S. (2002). Consensus treaties. *Journal of Institutional and Theoretical Economics*, 529–547.

Battaglini, M. and B. Harstad (2020). The political economy of weak treaties. *Journal of Political Economy 128*(2), 544–590.

Besley, T. and S. Coate (1998). Sources of inefficiency in a representative democracy: A dynamic analysis. *American Economic Review*, 139–156.

Buchholz, W., A. Haupt, and W. Peters (2005). International environmental agreements and strategic voting. *Scandinavian Journal of Economics 107*, 175–195.

Buisseret, P. and D. Bernhardt (2018). Reelection and renegotiation: International agreements in the shadow of the polls. *American Political Science Review 112*(4), 1016–1035.

Carothers, T. and A. O'Donohue (2019). *Democracies Divided: The Global Challenge of Political Polarization.* Brookings Institution Press.

Carraro, C. and D. Siniscalco (1993). Strategies for the international protection of the environment. *Journal of Public Economics 52*, 309–328.

Cazals, A. and A. Sauquet (2015). How do elections affect international cooperation? evidence from environmental treaty participation. *Public Choice 162*(3-4), 263–285.

Esteban, J. and G. Schneider (2008). Polarization and conflict: Theoretical and empirical issues. *Journal of Peace Research 45*(2), 131–141.

Finus, M. (2008). Game theoretic research on the design of international environmental agreements: Insights, critical remarks, and future challenges. *International Review of Environmental and Resource Economics 2*(1), 29–67.

Finus, M. and S. Maus (2008). Modesty may pay. *Journal of Public Economic Theory 10*, 801–826.

Gelman, A. and G. King (1990). Estimating incumbency advantage without bias. *American Journal of Political Science*, 1142–1164.

Hagen, A., J.-C. Altamirano-Cabrera, and H.-P. Weikard (2021). National political pressure groups and the stability of international environmental agreements. *International Environmental Agreements: Politics, Law and Economics 21*(3), 405–425.

Hänsel, M. C., M. A. Drupp, D. J. Johansson, F. Nesje, C. Azar, M. C. Freeman, B. Groom, and T. Sterner (2020). Climate economics support for the un climate targets. *Nature Climate Change 10*(8), 781–789.

Harstad, B. (2022). Pledge-and-review bargaining: From kyoto to paris. *The Economic Journal*.

Hoel, M. (1992). International environment conventions: The case of uniform reductions of emissions. *Environmental and Resource Economics 2*, 141–159.

Köke, S. and A. Lange (2017). Negotiating environmental agreements under ratification constraints. *Journal of Environmental Economics and Management 83*, 90–106.

Kolstad, C. D. and M. Toman (2005). The economics of climate policy. *Handbook of Environmental Economics 3*, 1561–1618.

Levin, S. A., H. V. Milner, and C. Perrings (2021). The dynamics of political polarization. *Proceedings of the National Academy of Sciences 118*(50), e2116950118.

Levitt, S. D. and C. D. Wolfram (1997). Decomposing the sources of incumbency advantage in the us house. *Legislative Studies Quarterly*, 45–60.

List, J. A. and D. M. Sturm (2006). How elections matter: Theory and evidence from environmental policy. *The Quarterly Journal of Economics 121*(4), 1249–1281.

Loeper, A. (2017). Cross-border externalities and cooperation among representative democracies. *European Economic Review 91*, 180–208.

Marchiori, C., S. Dietz, and A. Tavoni (2017). Domestic politics and the formation of international environmental agreements. *Journal of Environmental Economics and Management 81*, 115–131.

Melnick, J. and A. Smith (2022). International negotiations in the shadow of elections. *Journal of Conflict Resolution*, 00220027221139433.

Perrings, C., M. Hechter, and R. Mamada (2021). National polarization and international agreements. *Proceedings of the National Academy of Sciences 118*(50), e2102145118.

Persson, T. and G. Tabellini (1992). The politics of 1992: Fiscal policy and european integration. *The Review of Economic Studies 59*(4), 689–701.

Persson, T. and G. Tabellini (1994). Representative democracy and capital taxation. *Journal of Public Economics 55*(1), 53–70.

Putnam, R. D. (1988). Diplomacy and domestic politics: The logic of two-level games. *International Organization 42*, 427–460.

Robinson, J. A. and R. Torvik (2005). White elephants. *Journal of Public Economics 89*(2-

3), 197–210.

Siqueira, K. (2003). International externalities, strategic interaction, and domestic politics. *Journal of Environmental Economics and Management 45*(3), 674–691.

Spycher, S. and R. Winkler (2022). Strategic delegation in the formation of modest international environmental agreements. *European Economic Review*, 103963.

Swiss Re (2021). The economics of climate change: No action not an option. *Swiss Re Management Institute 30*, 50.

UNEP (2023). Emissions gap report 2023: Broken record – temperatures hit new highs, yet world fails to cut emissions (again). *doi.org/10.59117/20.500.11822/43922*.

Wagner, U. J. (2001). The design of stable international environmental agreements: Economic theory and political economy. *Journal of Economic Surveys 15*, 377–411.

Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

Strada Maggiore 45
40125 Bologna - Italy
Tel. +39 051 2092604
Fax +39 051 2092664
http://www.dse.unibo.it