Shaming Paris: A Political Economy of Climate Commitments*

Justin Melnick† and Alastair Smith‡

November 15, 2024

Abstract

We explore leaders' incentives to set climate commitments and subsequently exert downstream mitigation effort with a formal model. Since the Paris Agreement asks countries to make unilateral voluntary commitments, we investigate the domestic factors motivating climate pledges. We study a country with electoral competition between two parties, Green and Brown, who first make commitments to reduce emissions and then implement policies to meet their commitments. Voters anticipate the equilibrium policies each party will implement given the pledge. If downstream mitigation policies are insufficient relative to the commitment, the government is "shamed" by the international community. Several incentive channels arise when parties make commitments, as they have policy and electoral value. Parties can use commitments to tie the opposition's hands to implement preferential policies in the future. If parties solely care about winning elections, parties exploit commitments to serve electoral needs, which paradoxically leads anti-environmental parties to implement more ambitious commitments.

Word Count: 8999 (text) + 1466 (references) + 2328 (appendix) = 12793 (total)

*We thank audiences at New York University and the Jeju Peace Institute for insightful comments.

†Ph.D. Candidate, Department of Politics, New York University. Contact: melnickj@nyu.edu and justinmelnick.github.io.

‡Bernhardt Denmark Chair of International Relations, Department of Politics, New York University. Contact: as183@nyu.edu.

The 2015 Paris Agreement aims to hold the increase in global average temperature to below 1.5-2°C relative to pre-industrial levels through national pledges to voluntarily reduce carbon emissions. As of 2021, 181 of 195 signatory states have submitted "nationally determined contributions" (NDCs) with the United Nations Framework Convention on Climate Change (UNFCCC). Early assessments of NDCs document substantial cross-country heterogeneity in the ambition and relative burden-sharing of mitigation commitments (Holz, Kartha and Athanasiou 2018; Robiou du Pont and Meinshausen 2018); however, few studies attempt to explain this variation (Tørstad, Sælen and Bøyum 2020). This paper explores leaders' incentives to set climate commitments and subsequently exert costly effort toward implementing those commitments.

Leaders' decisions to make climate commitments and ultimately see them through are undoubtedly complex. Consider the NDC of the United Kingdom, which in December 2020 committed the country to reducing economy-wide greenhouse gas emissions by at least 68% by 2030, compared to 1990 levels.[1] This commitment was made by then-Prime Minister and member of the Conservative Party Boris Johnson, who claimed that "the UK will be home to pioneering businesses, new technologies and green innovation as we make progress to net zero emissions, laying the foundations for decades of economic growth in a way that creates thousands of jobs."[2] Indeed, Johnson's ambition was a watershed moment for the world, as the UK became the first major economy to pass a net zero emissions law.[3]

However, less than three years later, incumbent Tory Prime Minister Rishi Sunak announced he would push back the deadline for selling new petrol and diesel cars and the phasing out of gas boilers, key policy considerations in meeting the net zero target.[4] Sunak's rollback was scrutinized as an attempt to offload the green transition's anticipated costs away from voters in the run-up to the country's next election. The Labour Party was perceived as the greener party, liable to implement the costly policies to which the Conservatives had initially committed the country.[5] Sunak had hoped to pit the Conservatives as a cheaper, albeit less green, alternative to consumers than a potential Labour government. Of course, in the July 2024 election, issues other than climate change dominated the electoral landscape: the Conservatives' ability to make the issue salient on the margin proved insufficient and Labour won despite the perceived costliness of its carbon reduction policies.

[1] "United Kingdom of Great Britain and Northern Ireland's Nationally Determined Contribution," UNFCCC, September 2022. https://tinyurl.com/yr9bd4r5

[2] "UK enshrines new target in law to slash emissions by 78% by 2035," government press release, 20 April 2021. https://tinyurl.com/yc532fxc

[3] "UK becomes first major economy to pass net zero emissions law," government press release, 27 June 2019. https://tinyurl.com/zkkbrdtr

[4] "Rishi Sunak announces U-turn on key green targets," *The Guardian*, 20 September 2023. https://tinyurl.com/rfc579jw

[5] "Sunak U-turn on green policies forces Labour to revise its own," *The Guardian*, 20 September 2023. https://tinyurl.com/4pdyfvwj

We present a formal model probing the domestic motivations behind the setting of climate targets. Two parties, Green and Brown, compete for political office. Parties vary in their marginal costs of implementing mitigation policy: the Green party faces lower costs than the Brown party and is willing to commit to more expansive climate reforms *ex ante*. As the incumbent, one of these parties sets a national climate commitment; the party that wins the election then determines whether they want to implement domestic policy to meet that target. Voters support either the Green or the Brown party given the downstream mitigation measures induced by the nation's pledge. At the end of the game, nations imperfectly observe each leader's mitigation efforts and assess whether targets were met in a "global stocktake;" leaders deemed to have underdelivered relative to their national commitments are "shamed."

Our model is consistent with three central features of the Paris Agreement's structure. First is the notion that Paris seeks to recenter domestic politics into the implementation of international climate goals (Falkner 2016). Leaders choose their own commitments rather than accepting the terms of legally binding reduction targets, as with the Kyoto Protocol (Keohane and Oppenheimer 2016), which makes understanding the domestic political considerations surrounding the setting of commitments paramount. Second is the idea that leaders first determine NDCs and then must enact policies designed to fulfill the commitment made in the pledge. In the model, the incumbent party sets a commitment and then costly effort into mitigation strategies to meet the target is exerted in the future. Third is that these goals are not enforced formally; climate laggards incur reputational costs known as "naming and shaming" for failing to comply with their targets (Bodansky 2016; Jacquet and Jamieson 2016). We consider how the domestic electoral competition between political parties entices leaders to set different commitments in the shadow of possible international shaming.

Our analysis uncovers two relevant mechanisms through which domestic politics affect climate commitments. First, leaders may tie the hands of the opposition party and pick commitments to bring downstream policy measures closer to their preferred outcomes. Even devoid of electoral considerations, leaders may care about implementing climate goals on pure policy grounds. Leaders have policy preferences over possible levels of effort and can tailor their pledges to ensure that their preferred policies are implemented in the future. In particular, when elections are relatively insensitive to climate policy, the Green party can tie the hands of the Brown party after the election with an ambitious target. That is, the Green party can design their pledge in order to force the Brown party to enact policy closer to the Green party's policy preference.

A second set of incentives relates to the value of office-holding. If winning elections is leaders' dominant consideration, then they set commitments in order to maximize their electoral prospects based on the antic-

ipated costs of downstream mitigation measures to voters. The Green party faces an electoral disadvantage against the Brown party in this regard because the Green party would *ex ante* prefer to implement a more ambitious mitigation strategy than the median voter. The Brown party can leverage this advantage by counterintuitively embracing a lofty climate commitment. If the cost of being shamed is sufficiently large, the Green party will try to fulfill more ambitious pledges while the Brown party will not. The Brown party chooses an ambitious target, knowing it would not comply and would be shamed. However, the Green party would be willing to exert costly mitigation efforts in order to meet the goal, and this makes the Green party electorally unattractive to the median voter. Interestingly, while the commitment itself is uniform across national parties, voters' *expectations* about each party's likelihood of meeting it are different, which generates electoral incentives to exploit Paris's structure by anti-environmental parties (cf. Bagashka and Stone 2013). Ambitious climate commitments can therefore be leveraged in order to maximize electoral prospects based on how those commitments chart future national implementation measures and their subsequent costliness for voters.

By unpacking these mechanisms, our theoretical analysis explains how variation in observed climate commitments and subsequent policy outcomes arises. Given the Paris Agreement's structure, one country's commitment has no *direct* effect on the commitments of other countries. There is no reciprocity baked into the agreement's terms. Moreover, there is no international infrastructure to render these pledges "credible." Hence, heterogeneity in pledges and efforts to implement these pledges are driven by variation from politics within nations. Our model points to changes in domestic fundamentals – like variation in the median voter's willingness to pay for climate policies, public support for climate policy as an electoral issue, parties' valuation of holding office, and parties' valuations for climate policy, among other parameters – which point to different incentives that may drive leaders to implement more or less ambitious climate commitments. It is these domestic forces that also ultimately guide the extent to which leaders see these commitments through.

We contribute broadly to the literature on the domestic and international political economy of climate agreements. Much of the recent work in climate politics focuses on public opinion (Gazmararian, Mildenberger and Tingley 2024). Experimental work consistently finds that individual support for climate policy and politicians advocating such policies are highly contingent upon the expected costs (e.g., Bechtel and Scheve 2013; Ansolabehere and Konisky 2014; Gazmararian and Tingley 2023). Scholars have sought to identify consumers' willingness to pay for particular climate policies (e.g., Nemet and Johnson 2010; Kotchen, Boyle and Leiserowitz 2013) and whether there exist broad "climate coalitions" in favor of climate-friendly policies

(e.g., Bergquist, Mildenberger and Stokes 2020; Gaikwad, Genovese and Tingley 2022). We complement this work in two ways. First, we provide a theoretical rationale for leaders' politically optimal climate policies in the shadow of domestic support, which assists in explaining the intensity of mitigation policy that should be expected in equilibrium. Second, we demonstrate how leaders internalize voters' anticipated costs of implementing climate policy in setting their nationally determined contributions *ex ante* and how these costs may be leveraged for electoral gain.

Theoretically, our model fits squarely within the "two-level games" tradition of modeling international cooperation (Putnam 1988; Milner 1997). We characterize the effects of elections on the incentives to commit to international treaties (Buisseret and Bernhardt 2018; Melnick and Smith 2023). Battaglini and Harstad (2020) demonstrate leaders' electoral incentives to sign "weak treaties" in which leaders overcommit but may underdeliver on their environmental promises. Köke and Lange (2017) also consider the ratification of international environmental agreements from a domestic perspective and investigate the role of uncertain ratification on the depth of commitments. Dai (2007) also finds that governments exhibit greater compliance with international treaties when pro-compliance domestic groups have more electoral leverage and informational capacity using the case of the Convention on Long-range Transboundary Air Pollution. Additionally, as is common in two-level games, we highlight how the preferences of domestic actors may serve as an endogenous veto constraint on the ability to implement international commitments (Chapman, Urpelainen and Wolford 2013; Iida 1993; 1996; Milner and Rosendorff 1997; Mo 1995), here reflected in the voters' willingness to pay for mitigation measures.

Extant approaches to modeling the Paris Agreement have interrogated the effects of its novel institutional features on the prospects for climate cooperation. For example, Harstad (2023*b*) presents a dynamic bargaining model that documents the conditions under which the Paris Agreement yields more ambitious climate commitments than the Kyoto Protocol. Other models capture how Paris's role in disseminating information affects the scope for ambitious contributions (Harrison and Lagunoff 2017; Slechten 2020; McAllister and Schnakenberg 2022). While we are not the first to examine a formal model of climate change cooperation, ours is the first to provide a domestically-microfounded story of the implementation of the Paris Agreement that goes beyond global collective action concerns (Aklin and Mildenberger 2020; Kennard and Schnakenberg 2023).

The Paris Climate Accord has no inherent means of sanctioning noncompliance and, as we shall see, reputational costs will play a sizable role in determining equilibrium commitments. We therefore contribute to the literature interrogating the efficacy of naming and shaming (Hafner-Burton 2008; Terman 2023).

Problems like information transmission (Creamer and Simmons 2019; Raiser, Çalı and Flachsland 2022) or issue politicization (Terman and Voeten 2018) may stymie naming and shaming and thus weaken compliance, while strategies such as issue linkage (Hafner-Burton 2005; Spilker and Böhmelt 2013) may enhance reputational incentives to comply. Despite potential shortcomings, recent studies of policy elites demonstrate that policymakers view naming and shaming as an adequate, and even preferable, means of sustaining cooperation (Hafner-Burton, LeVeck and Victor 2017; Dannenberg et al. 2023).

How naming and shaming "works" is central to cooperation under Paris and therefore highly relevant in our study. While we follow the human rights literature and think about shaming coming from international actors in a reduced-form way (Hafner-Burton 2012), empirical studies have also sought to tease out domestic microfoundations for compliance. Tingley and Tomz (2022) find that shaming by other countries increases support for climate commitments. Other experimental work also demonstrates the presence of shaming costs for leaders who fail to live up to their promises (Casler, Clark and Zucker 2023; Andrews and Bokemper 2024).

Finally, we complement a burgeoning empirical literature on the effects of the Paris Agreement and the determinants of NDCs. Tørstad and Wiborg (2023) use a conjoint experiment to demonstrate that the likelihood of compliance is a strong determinant of general public support for climate agreements. In general, empirical evidence suggests that the quality of national political institutions explains most variability in "credible" climate commitments (Victor, Lumkowsky and Dannenberg 2022). Wealthier countries pledge to undertake greater emission reductions with higher costs (Aldy et al. 2016), and more democratic countries and countries more vulnerable to climate change have been associated with more ambitious commitments (Tørstad, Sælen and Bøyum 2020). However, given the complexity in setting policy to meet mitigation targets, some scholars have argued that it is difficult to know if Paris targets across countries are empirically comparable (Rowan 2019). Hence we provide a theoretical treatment of NDCs and the domestic political forces that shape them.

Paris and Climate Commitments

The Paris Agreement seeks to overcome the global collective action problem by encouraging voluntary emissions reduction commitments enforced through reputational sanctions. Article 4.2 of the Agreement requires that "Each Party shall prepare, communicate and maintain successive nationally determined contributions that it intends to achieve" (UNFCCC 2015). Rather than delegate authority to an international body

that imposes top-down, legally binding targets as in other international climate governance frameworks like the Kyoto Protocol, nations asymmetrically consider their own incentives and abilities to abate (Harstad 2023*a*;*b*). As negotiated, such an institutional design is maximally "flexible" (Rosendorff and Milner 2001; Johns 2014; Linos and Pegram 2016), albeit very "shallow" (Downs, Rocke and Barsoom 1998; Gilligan 2004; Edry 2020). Taking these institutional attributes as exogenous, we consider the domestic political incentives to make commitments within such an agreement. Important for our story, these initial commitments serve as *endogenous reference points:* climate pledges, while chosen strategically, may redefine the scope of desirable policies that leaders implement in the future (cf. Leinaweaver and Thomson 2021).

The Paris Agreement does not explicitly identify any enforcement mechanism to ensure that nationally determined contributions are implemented. Article 7.14 establishes the system of pledge-and-review in which nations reconvene for a "global stocktake" to assess progress toward achieving NDCs and inform future measures (UNFCCC 2015). Articles 13, 14, and 15 outline the "enhanced transparency framework" and information dissemination process intended to serve as compliance mechanisms. The first global stocktake occurred in 2023, and they are set to be held every five years thereafter. As part of the process, countries submit reports about their performance (a noisy signal of their effort into meeting their pledges). The stocktake itself does not serve to "name and shame" individual countries (Milkoreit and Haapala 2017), but it does generate a report laying the groundwork for assessments of policy goals by other nations and public actors like NGOs and activists to pressure leaders into adopting more ambitious commitments (Hermwille and Kreibich 2018). This process thus provides a platform to facilitate international shaming, deemed effective by policymaking elites (Dannenberg et al. 2023).

Given the long time horizon between submission of NDCs and subsequent evaluations of progress, enforcement of the agreement is thus informal and if costs of noncompliance are imposed they are levied in the future, not when nations initially set their targets. Leaders who set their nation's commitments need not be in power when it comes time to "name and shame" those who did not follow through on their commitments.

Since pledges are not legally binding and enforcement is uncertain, leaders vary in their ultimate willingness to comply with their nation's target. We stipulate that leaders pay a "shaming cost" if judged to have failed to fulfill their commitment. This cost is larger if leaders anticipate greater reputational sanction for breaching their commitment, and larger expectations of shaming costs can entice leaders to fulfill larger commitments. However, in a world with imperfect monitoring (Porter 1983; Dai 2002), the precision with which the international community can verify national emissions reductions also affects leaders' incentives to comply with the target. As we will demonstrate, downstream mitigation efforts are dependent on the

interaction between these international factors and domestic policy preferences.

Model Setup

Our model illustrates a multistage policy process in which nations gather at a multilateral summit to pledge emissions reductions and subsequently enact policies to meet those reduction targets. There are n countries indexed by $i = 1, \ldots, n$. We will focus on the decision-making of a representative nation that is governed by one of two governments, $g \in \{G,\ B\}$ (and omit subscript i where it is not confusing). Governments vary in their marginal costs of implementing emissions reductions, λ_g. A "Green" government G faces lower marginal costs than a "Brown" government B, so $\lambda_G < \lambda_B$. The nation also includes a median voter M such that $\lambda_G < \lambda_M < \lambda_B$.

Each country initially sets their target $y \in \mathbb{R}_+$, which is analogous to their nationally determined contribution in the Paris framework. This represents the overall reduction in carbon emissions to be achieved by the nation by the end of the pledge-and-review period. After setting their targets, nations implement mitigation strategies and other policy measures designed to meet their targets, $a \in \mathbb{R}_+$. We endow actors of our representative nation with the following utility function over policy:

$$u_g(a_g, A; \lambda_g) = A - \frac{\lambda_g}{2} a_g^2, \qquad g \in \{B, G, M\},$$

where $A = \sum_i a_i$ is global emissions reductions. We suppress dependence on A and λ_g where it is not confusing, writing $u_g(a)$.

All nations benefit when others enact policies to reduce emissions, hence utility is increasing in the mitigation efforts of other countries, but mitigation is costly at home. Pursuing more ambitious reductions yields increasing marginal costs, as reflected by the quadratic term, with λ_g parameterizing the magnitude of these marginal costs. In what follows, it will be convenient to denote the reduction target that maximizes this function as actor g's "ideal point," $\tilde{a}_g = \frac{1}{\lambda_g}$.

After nations set their targets but prior to the implementation of mitigation policy, there is an election in our representative nation. We place the election in between these two points of the game in order to study the electoral incentives to enact different *commitments*, which, as we shall see, will indirectly affect their choices of mitigation policy as well. The election is determined by the median voter M, who incurs costs to adjust to mitigation strategies such that $\lambda_G < \lambda_M < \lambda_B$. That is, the median voter prefers greater emissions reductions than the Brown government, but does not share the ambition of the Green government. For the

purposes of constructing examples, we let $\lambda_M = \frac{\lambda_B + \lambda_G}{2}$. The median voter is prospective, and votes for the Green government if and only if the payoff from electing the Green government exceeds that from electing the Brown government. In addition to observing the pledge y, the median voter observes valence shocks μ_G and μ_B that represent the value of both parties on all other electorally salient dimensions beyond mitigation policy. Let $\mu = \mu_B - \mu_G$ such that $\mu \sim F(\cdot)$ with associated density $F'(\cdot)$. Thus, the median voter prefers the Green government if and only if

$$u_M(a_G) - u_M(a_B) \geq \mu,$$

and so G's probability of election is $F\left(u_M(a_G) - u_M(a_B)\right)$.

Finally, as the pledge-and-review period ends, nations reconvene for a "global stocktake" that examines how successful countries were in implementing their targets. This amounts to determining the distance between a and y. We assume that each a is imperfectly observed: the international community observes a noisy signal of the reduction measures $x = a + \varepsilon$, where $\varepsilon \sim N(0, \frac{1}{\beta})$. If it is determined that country i failed to reach its target, which occurs if $x < y$, then the governing party in country i is "shamed" and incurs a cost $\sigma_g \in \mathbb{R}_+$. We allow the impact of shaming to vary across parties. Although not necessary for our results, we may anticipate that the Green party faces a larger cost for failing to follow through on its commitment than the Brown party. Thus, given a commitment of y and effort level a the ruling party is shamed with probability $\Phi(\sqrt{\beta}(y - a))$ where $\Phi(\cdot)$ and $\phi(\cdot)$ are the cumulative distribution and probability density functions for the standard normal respectively. Observe that, while the magnitude of the shaming cost is fixed regardless of the size of the transgression between x and y, leaders have varying expectations about the probability they will be shamed as a function of how much effort they exert into fulfilling the commitment.

While the framers of Paris had hoped that naming and shaming could originate from international and domestic sources (Falkner 2016), our preferred interpretation is that the shaming cost is reputational and levied by other nations on noncompliant states. This is different from an endogenous cost levied upon leaders by voters, but this strategic dynamic is explored elsewhere in the literature (e.g., McGillivray and Smith 2008).[6] An external source of shaming also comports with experimental evidence showing individuals are more likely to support commitments if they know their leaders could be shamed (Tingley and Tomz 2022). Hence, the shaming cost σ_g should be thought of as conceptually distinct from the median voter's decision

[6]To be clear, we do not interpret σ_g as an "audience cost" in the sense of Fearon (1994), although see Casler, Clark and Zucker (2023) for such an interpretation.

to retain or replace the incumbent party given the observed commitment (prior to implementation). We understand that there may be credibility or collective action problems in terms of who does the shaming internationally (Hafner-Burton 2008; Terman 2023), but the probabilistic nature of shaming in our model captures these concerns in reduced-form.

Finally, let $\rho \in \{0, 1\}$ denote whether the median voter elects the Green party ($\rho = 1$) or the Brown party ($\rho = 0$). Governments thus have the following payoff from making a commitment y,

$$v_G(y) = \rho(\Psi + u_G(a_G) - \sigma_G \Phi(\sqrt{\beta}(y - a_G))) + (1 - \rho)(u_G(a_B)).$$
$$v_B(y) = \rho(u_B(a_G)) + (1 - \rho)(\Psi + u_B(a_B) - \sigma_B \Phi(\sqrt{\beta}(y - a_B))).$$

This payoff demonstrates that, when choosing climate commitments, leaders care about mitigation policy outcomes, the ability to influence electoral outcomes through the behavior of the median voter, and winning elections. The party that wins the election enjoys benefit $\Psi > 0$. Notice also that only the party in power incurs the shaming cost σ_g if their mitigation efforts are judged to fall short of the nation's climate commitment. We do not require that the median voter nor the party out of power pays the shaming cost, although many of the main features of the equilibrium would be robust to this modification.

The timing of the game is summarized as follows:

1. Governments commit to pledges y.
2. The median voter observes their nation's pledge y and votes to elect either the Green government G or the Brown government B.
3. The elected government implements mitigation policies a_g.
4. Nations review global mitigation progress and observe x, shaming country i if $x < y$.

We analyze the subgame perfect equilibrium. The incumbent party chooses a climate commitment $y_g \in \mathbb{R}_+$. The median voter's strategy is a mapping from the expected efforts given y and the valence shock into a vote choice for G or B. Finally, the party that wins the election chooses effort $a_g \in \mathbb{R}_+$ given their nation's prior commitment.

Analysis

We start with a general characterization of the subgame perfect equilibrium. Using backward induction we characterize each party's mitigation efforts for each possible commitment, find how the efforts affect the voter's electoral decision, and characterize the commitments that each party will make given how such commitments affect the election and subsequent mitigation efforts. In equilibrium, a party's climate pledge is influenced by a variety of factors including the ability to tie the other party's hands with respect to policy implementation and influencing which party will win election. To isolate the properties of each of these mechanisms, we examine a series of limiting cases as the signals of mitigation efforts become precise ($\beta \to \infty$).

Optimal Mitigation Efforts

We first consider the emissions reduction target pursued by government g after the election. Government g's expected utility is

$$u_g(a_g, A; \lambda_g) = \underbrace{A}_{\text{benefits}} - \underbrace{\frac{\lambda_g}{2} a_g^2}_{\text{cost of mitigation}} - \underbrace{\Phi(\sqrt{\beta}(y - a_g))\sigma_g}_{\text{expected shaming}},$$

which, given the predetermined pledge y, is the utility over mitigation commitments plus the probability of being shamed and incurring the cost σ_g for failing to meet the pledge. The optimal mitigation effort a_g^* therefore solves the following first-order condition stated in Lemma 1.

Lemma 1 *Given climate commitment y, the government's policy a_g^* satisfies the first-order condition (FOC)*

$$\frac{du_g(a_g, A; \lambda_g)}{da_g} = 1 - \lambda_g a_g^* + \sigma_g \sqrt{\beta} \phi(\sqrt{\beta}(y - a_g^*)) = 0 \tag{1}$$

and the second-order condition (SOC)

$$\frac{d^2 u_g(a_g, A; \lambda_g)}{da_g^2} = -\lambda_g + \sigma_g \beta \sqrt{\beta}(y - a_g^*)\phi(\sqrt{\beta}(y - a_g^*)) < 0. \tag{2}$$

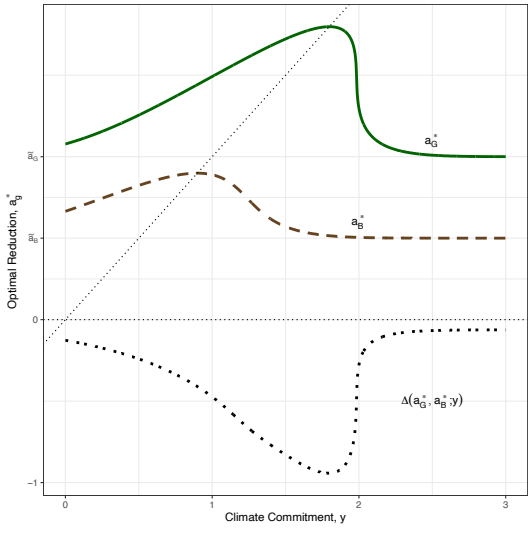
If the signal of effort x is sufficiently noisy, there is a unique solution to the FOC (equation 1). However, with precise signals, there might be two local maxima that satisfy the FOC which can result in a discontinuity in the government's optimal response. Given the technical rather than the substantive nature of this uniqueness discussion, we characterize these conditions at length in the appendix (Lemmas A.1 and A.2).

Parties weigh the marginal costs of exerting effort with their global marginal benefits and the possibility of being shamed and, in general, will do one of two things: they will either set effort close to their ideal point $\tilde{a}_g$ or they will set effort close to the pledge y. Leaders will choose effort close to their ideal point if the commitment is sufficiently low such that exerting effort close to their ideal point is sufficient to avoid being shamed, or if the commitment is extremely high such that parties do not find it in their interest to pursue it and rather prefer to accept that they are likely to be shamed. By contrast, if the commitment is not too high relative to the incumbent's ideal point, then governments will exert effort closer to the target in order to lower the probability of being shamed.

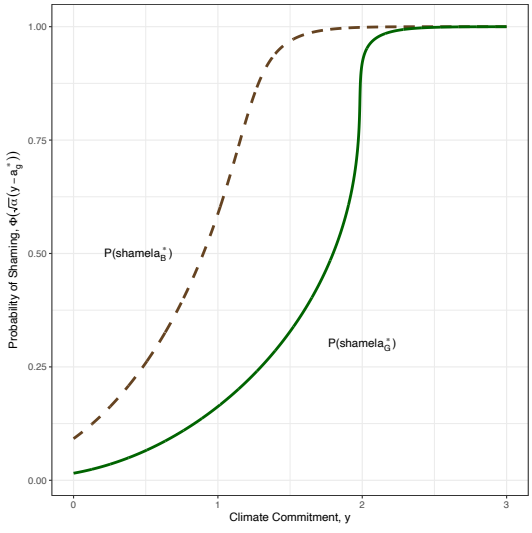
Figure 1 plots the optimal efforts of each party and the likelihood of being shamed as a function of the climate commitment y.[7] The left panel plots G's optimal effort in green (solid line) and B's optimal effort in brown (dashed line). Optimal efforts are non-monotonic in the commitment y. If y is not too large then parties may find it in their interest to comply with the target as a means of avoiding shame. However, this incentive dissipates if y is set too ambitiously, as the costs of effort to avoid being shamed grow; parties resign themselves to being shamed and revert effort close to their ideal point. The Green party exerts greater effort than the Brown party in equilibrium, because their marginal costs of exerting effort are smaller, and exerting greater effort means that G will be shamed with a smaller probability than B, illustrated in the right panel of the figure.

Figure 1: Efforts and the Likelihood of Shaming as a Function of Commitments

(a) Optimal Efforts and Electoral Bias



(b) Probability of Shaming



[7]The figure is constructed using the parameters $\lambda_G = 1,\ \lambda_B = 2,\ \beta = 4,\ \sigma_G = \sigma_B = 1$.

Leaders' incentives to comply *ex post* with international mitigation targets thus depend on how ambitiously the commitment was set relative to their ideal effort level and the chances that they could be shamed for noncompliance. Of course, the government in power when commitments were set need not be the government tasked with implementing policy to meet those commitments: this depends on how voters perceive the costs of future mitigation policies and the extent to which these policy concerns affect the electoral outcome.

Voting Behavior

We now consider the behavior of the median voter. When choosing whether to elect the Green government or the Brown government, M anticipates the mitigation efforts that each party will make and how costly these policies will be for her. Empirical work has reflected that voters' willingness to support mitigation policy is highly sensitive to the costs of those policies (e.g., Bechtel and Scheve 2013; Ansolabehere and Konisky 2014; Gaikwad, Genovese and Tingley 2022); the median voter's electoral decision reflects these sensitivities. Moreover, we acknowledge that the salience of climate policy may be low to voters – although increasing over time (Egan, Konisky and Mullin 2022) – so the median voter also evaluates the two possible governments along other electorally relevant considerations, captured by the valence terms μ_g. Quite simply, M votes for the Green government over the Brown government when

$$u_M(a_G^*) + \mu_G \geq u_M(a_B^*) + \mu_B \;\Leftrightarrow\; \mu \leq u_M(a_G^*) - u_M(a_B^*) \;\equiv\; \Delta(a_G^*, a_B^*; y).$$

Climate commitments affect voting outcomes through their expected costs on the median voter in implementing effort needed to fulfill those commitments. For any y, parties implement their optimal a_g^* after the election, which the voter can anticipate. This means that, in the model, the voter observes the commitment prior to the election, and forms an expectation of the policies that each party would implement should they come to office. This is not the same as punishing a party who failed to meet their commitment. Instead, the voter adjudicates the relative costliness of G and B's expected policies against other electorally salient issues. Straightforwardly, as the Green government proposes more and more ambitious commitments relative to the Brown government, the median voter's expected costs from voting for the Green government increase, which makes the Green party less attractive electorally. We term the difference in M's policy utility from G and B's equilibrium efforts as the "bias" toward the Green party denoted Δ; the Green party is therefore elected with probability $F(\Delta)$.

Optimal Climate Commitments

We now turn to the optimal pledges that different governments would set. Leaders care directly about the policy returns from committing to a pledge y, as well as holding office. These concerns in turn affect the median voter's willingness to reelect incumbents based on the prospective mitigation policies to be chosen after the election. The choice of climate commitment affects party G's payoff as follows:

$$V_G(y) = F(\Delta)(\Psi + u_G(a_G^*) - \sigma_G \Phi(\sqrt{\beta}(y - a_G^*))) + (1 - F(\Delta))u_G(a_B^*).$$

With probability $F(\Delta)$, G wins the election, gets office benefits Ψ, and implements effort a_G^*, knowing that with probability $\Phi(\sqrt{\beta}(y - a_G^*))$ they will be shamed. However, with probability $1 - F(\Delta)$, party B wins the election and G receives the policy payoff associated with B's equilibrium effort.

Likewise party B's payoff is

$$V_B(y) = F(\Delta)u_B(a_G^*) + (1 - F(\Delta))(\Psi + u_B(a_B^*) - \sigma_B \Phi(\sqrt{\beta}(y - a_B^*))).$$

We write y_g^* to be the optimal climate commitment that party g chooses, maximizing their payoff,

$$y_g^* \in \arg\max_{y \in \mathbb{R}_+} V_g(y).$$

Given our backward induction analysis, we can now straightforwardly summarize the preceding discussion of optimal effort, voting decisions, and selection of climate commitments.

Proposition 1 *In subgame perfect equilibria, party G selects y_G^* and implements effort a_G^* if elected; party B selects y_B^* and implements effort a_B^* if elected; the median votes for G if and only if $\mu \leq \Delta(a_G^*, a_B^*; y)$ and G is elected with probability $F(\Delta(a_G^*, a_B^*; y))$.*

A party's choice of the commitment y will affect several factors that are driven by how pledges shape downstream mitigation efforts after the election. Climate commitments can thus be useful in policy terms, as parties may be able to tie the hands of their competitors through their choice of pledge. Moreover, because pledges affect effort levels, they affect who wins the election. This latter factor is encapsulated through the commitment's effect on Δ, the net electoral value of G relative to B. In such a general setting, it is difficult to isolate the substantive impact of these competing policy and office incentives. As explored in the appendix, party payoffs may look very different depending on which incentives dominate. To isolate the influence of

each factor on climate commitments, we look at a series of limiting cases. The value of examining these limiting cases is also to demonstrate how variation in primitives isolates different concerns for governments when shaping their climate commitments, which ultimately drives variation in observed outcomes.

Limiting Case: Precise Shaming

We now present a special case of our model in which the uncertainty around shaming vanishes, $\beta \to \infty$, meaning leaders know whether they will be shamed with certainty. Optimal efforts are fairly simple in this case: the election winner will either comply with the target or will implement their ideal effort level. If the target y is low, then leaders can implement their ideal point and avoid shaming. Increasing the ambition of the climate commitment ($y > \tilde{a}_g$) means leaders face a trade-off between implementing the target or implementing their ideal point instead and incurring the costs of being shamed. If a party complies with the pre-existing target, their payoff is $y_g - \frac{\lambda_g}{2} y_g^2$ where the first term corresponds to the benefits of implementing the target and the quadratic term represents the costs. Alternatively, the party might implement its ideal point and get shamed, which yields a payoff of $\tilde{a}_g - \frac{\lambda_g}{2}\tilde{a}_g^2 - \sigma_g$. Hence, whenever $y \leq \hat{y}_g = \frac{1+\sqrt{2\lambda_g \sigma_g}}{\lambda_g}$, leaders prefer to comply with the target instead of implementing their ideal point and getting shamed. If the pledge is set too ambitiously, $y > \hat{y}_g$, then leaders will revert to implementing their ideal level of effort, knowing they will be shamed. This is summarized in the following corollary.

Corollary 1 *Let $\beta \to \infty$. Government g pursues the mitigation effort*

$$a_g^*(y) = \begin{cases} \tilde{a}_g & \text{if } y < \tilde{a}_g \\ y & \text{if } \tilde{a}_g \leq y \leq \hat{y}_g \\ \tilde{a}_g & \text{if } y > \hat{y}_g. \end{cases}$$

Corollary 1 tells us the downstream policies that each party will implement if elected given pledge y. We now turn to thinking about the incentives that parties face when choosing the commitments themselves. Parties might pick a commitment in order to tie the hands of an opposition party to implement a policy they like. Alternatively, a party might pick a commitment to gain an electoral advantage. Using the precise shaming technology, we examine each of these mechanisms.

Tying Hands

Suppose first that parties choose climate commitments solely for their policy value. To isolate this mechanism, we assume that holding office is irrelevant, $\Psi \to 0$, and that elections are not sensitive to climate policy, $F' \to 0$. Setting a commitment is valuable insofar as it ties politicians' hands when enacting future mitigation efforts. Since climate commitments serve as endogenous reference points, they define the scope of possible policies that could be implemented in the future. Thus, choosing a commitment has value if incumbents can ensure that potential electoral opposition will not deviate from climate policies they like. Parties set commitments to affect the implementation of effort a_g^* after the election.

The incentive to tie hands is particularly important for G, who sets a commitment y that forces B to increase its climate investments, more than would be the case absent a commitment. Ideally, G would like B to implement G's ideal point, which relies on the possibility that B could be sufficiently shamed for failing to follow through on this policy. If the shaming effect σ_B is insufficient for G to force B to implement G's ideal point, then G sets the target to the largest policy that B would be willing to fulfill, should B come to power after the election. Formally, we define $\hat{\sigma} = \frac{(\lambda_B - \lambda_G)^2}{2\lambda_G^2 \lambda_B}$ as the smallest shaming cost such that party B would be willing to adhere to a climate commitment at G's ideal point, (i.e. $u_B(\tilde{a}_G) = u_B(\tilde{a}_B) - \hat{\sigma}$). For large costs of shaming, in particular $\sigma_B \geq \hat{\sigma}$, G can fully tie B's hands and force it to implement G's ideal point by choosing $y_G^* = \tilde{a}_G$. This ensures that effort will be set at G's ideal point, regardless of who wins the election. However, if shaming costs are lower ($\sigma_B < \hat{\sigma}$), G cannot induce B to exert effort at G's ideal point $\tilde{a}_G$; B would rather be shamed than implement such ambitious climate reforms. Instead, G ties B's hands to the greatest extent possible by setting $y_G^* = \hat{y}_B = \frac{1+\sqrt{2\lambda_B \sigma_B}}{\lambda_B}$. This pledge is the greatest y that B would be willing to comply with, making B indifferent between exerting effort at the pledge y, avoiding shaming, and implementing its ideal point $\tilde{a}_B$ and incurring the shaming cost σ_B.

By contrast, B cannot tie G's hands at all, as G ideally prefers to exert greater effort than B. The best that B can do is set a target at no more than its ideal point: this allows B to remain in compliance with the agreement should B win the election. If G wins the election, G would implement its own ideal point. Proposition 2 summarizes this discussion.

Proposition 2 *Let $\beta \to \infty$, $\Psi \to 0$, and $F' \to 0$. G's optimal commitment is*

$$y_G^* = \begin{cases} \hat{y}_B = \frac{1+\sqrt{2\lambda_B \sigma_B}}{\lambda_B} & \text{if } \sigma_B < \hat{\sigma} \\ \tilde{a}_G & \text{if } \sigma_B \geq \hat{\sigma}. \end{cases}$$

B's optimal commitment is any

$$y_B^* \leq \tilde{a}_B.$$

Figure 2: Climate Commitments with Tying Hands Incentives

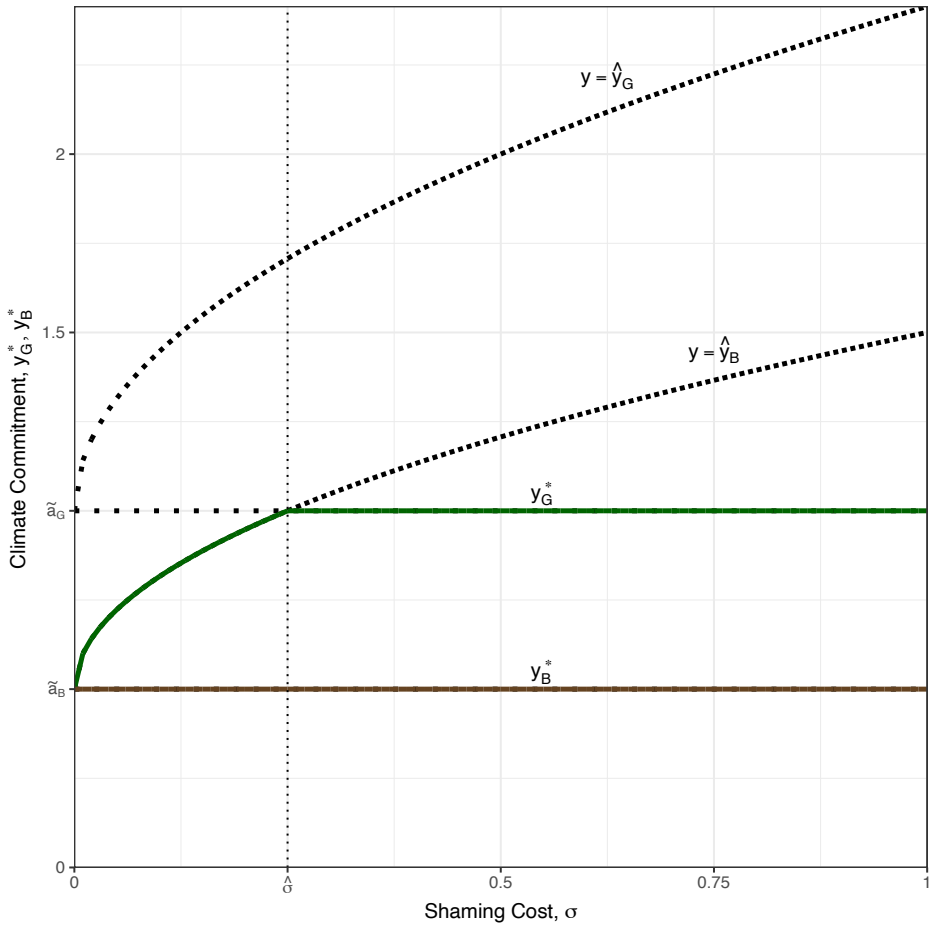


Figure 2 illustrates the optimal climate commitments as a function of the shaming cost σ_B. The green and brown lines plot each party's optimal commitment. If parties only care about the policy value of climate commitments, then the Green party can use climate pledges to drag the Brown party's effort on climate change as close to Green's ideal point as possible. Ideally, each party would like to implement their own ideal point, $y_g = \tilde{a}_g$, fully tying the hands of the other party. However, if σ_B is too small, G cannot completely tie B's hands; instead, G sets y_G^* such that B exerts the most effort it would be willing to invest prior to becoming shamed. This is seen on the left-hand side of Figure 2 as G's increasing optimal target when $\sigma_B \leq \hat{\sigma}$. B, who cannot tie G's hands, simply sets the most ambitious commitment that allows for the implementation of its ideal point.

Winning Office

We now examine how parties can use climate commitments to help them remain in elected office. To model these electoral incentives, we let the value of office-holding grow large such that winning office becomes the dominant incentive for parties, $\Psi \to \infty$. Choosing pledges therefore depends on maximizing the probability of winning the election; recall that the median voter votes for G when the valence μ is less than the difference in policy utility from each party's anticipated climate investments, $\mu \leq u_M(a_G^*) - u_M(a_B^*)$. This difference in utility $\Delta(a_G^*, a_B^*; y) = u_M(a_G^*) - u_M(a_B^*)$ is the "bias" that the voter has toward the Green party.

When choosing a climate commitment, G wants to maximize the electoral bias, while B wants to minimize it. Figure 3 plots the electoral bias $\Delta(a_G^*, a_B^*; y)$ as a function of the commitment y (for relatively large σ_g). An examination of how pledges affect the electoral bias facilitates the exposition of pledge-setting in the office-holding environment.

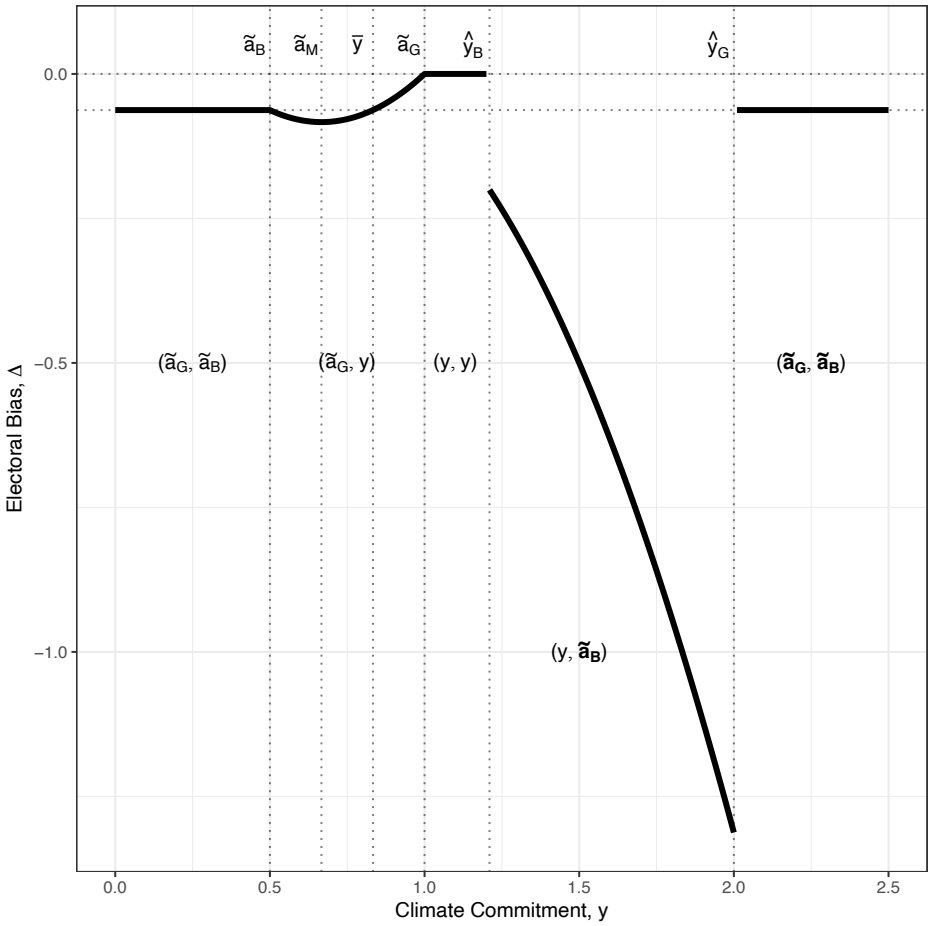
Climate commitments below $\tilde{a}_B$ provide no constraint on the implementation of climate policy; both parties would implement their ideal point if elected. The result is a small baseline electoral bias in B's favor, $\Delta(\tilde{a}_G, \tilde{a}_B; y) < 0$, shown by the flat line on the left of Figure 3.[8] Likewise, if the commitment is very large, $y > \hat{y}_G$, then neither party would attempt to fulfill the commitment; both parties would implement their ideal points and both would be shamed. Such an overly ambitious target results in the same baseline electoral bias in B's favor and is shown by the the flat line on the right of Figure 3.

Next consider the range of commitments between B and G's ideal points: $y \in (\tilde{a}_B, \tilde{a}_G)$. In this region, B would implement the target, while G would implement its ideal point. As the pledge initially increases above B's ideal point the electoral bias decreases (i.e. moves in B's favor) as the policy that B would implement moves closer to the median voter's ideal point (while G still implements its own ideal point). The electoral bias reaches a local minimum at $y = \tilde{a}_M$ when B implements the median voter's ideal point. As y further increases above $\tilde{a}_M$ the electoral bias $\Delta(\tilde{a}_G, y; y)$ increases (i.e moves in G's favor) as the policy B would implement if elected moves above the median voter's ideal point.

As y increases toward $\tilde{a}_G$ the electoral bias becomes zero. For pledges between $\tilde{a}_G$ and $\hat{y}_B$, both parties would implement the climate commitment so on the basis of climate policy there is no difference between the parties, $\Delta(y, y; y) = 0$. For the Green party, making a climate commitment of $y = \tilde{a}_G$ within this region is highly desirable as it removes the baseline electoral bias in favor of B while simultaneously tying B's hands

[8] Given our working assumption that $\lambda_M = \frac{\lambda_G + \lambda_B}{2}$, there is a slight electoral bias toward the Brown party if both parties are expected to enact their ideal points, $\Delta(\tilde{a}_G, \tilde{a}_B) = -\frac{(\lambda_B - \lambda_G)^3}{4\lambda_G^2 \lambda_B^2} < 0$. This bias emerges because the voter's utility function exhibits quadratic loss: it is more costly to move to a more ambitious policy in the direction of the Green party's ideal point than to move to a less ambitious policy in the direction of the Brown party's ideal point.

Figure 3: Climate Commitments and Electoral Bias



to implement G's ideal point.

Recall that $\hat{y}_g$ is the largest commitment that party g would adhere to before preferring to implement their ideal point even though this means being shamed. Above $\hat{y}_B$, the electoral bias jumps downward (i.e. in B's favor). For commitments in the range $y \in (\hat{y}_B, \hat{y}_G]$, G would implement the pledge if elected but B would implement its ideal point and be shamed. Since the voter prefers B's ideal point to the implementation of such large commitments – ambitious pledges that G is willing to implement, but B is not – these pledges push the election in B's favor; and the larger the pledge (subject to $y \leq \hat{y}_G$) the more it helps B electorally. Paradoxically, the Brown party has an electoral incentive to make bold pledges knowing that they will not carry them out but knowing the Green party would.

If parties care primarily about office holding, then G pledges a commitment that maximizes the electoral bias, while B wants to minimize it. The above analysis of Figure 3 provides a simple characterization of

G and B's commitments provided the cost of being shamed is sufficiently large. Green would pick its ideal point, $y_G^* = \tilde{a}_G$, as this maximizes the electoral bias. Brown would pick $y_B^* = \hat{y}_G$ that minimizes the electoral bias. However, such a analysis is valid only when σ_g is sufficiently large. When the cost of being shamed is relatively small for either party, such large commitments cannot be credibly implemented.

The proposition below specifies the optimal commitments for office-seeking parties for all possible shaming costs. For ease of exposition we focus on the special case where both parties face the same shaming cost, $\sigma_B = \sigma_G = \sigma$. In the Appendix, Proposition A.1 relaxes this assumption and finds largely identical behavior.

Proposition 3 *Let $\beta \to \infty$, $\Psi \to \infty$, $F' > 0$ and $\sigma_B = \sigma_G = \sigma$. There exist thresholds $\bar{\sigma}$, $\hat{\sigma}$, and $\hat{\hat{\sigma}}$ such that G's optimal climate commitment is*

$$y_G^* = \begin{cases} y \leq \tilde{a}_B \text{ or } y \in (\hat{y}_B, \tilde{a}_G] & \text{if } \sigma_B < \bar{\sigma} \\ \hat{y}_B = \frac{1+\sqrt{2\sigma_B \lambda_B}}{\lambda_B} & \text{if } \bar{\sigma} \leq \sigma_B \leq \hat{\sigma} \\ \tilde{a}_G & \text{if } \sigma_B > \hat{\sigma}. \end{cases}$$

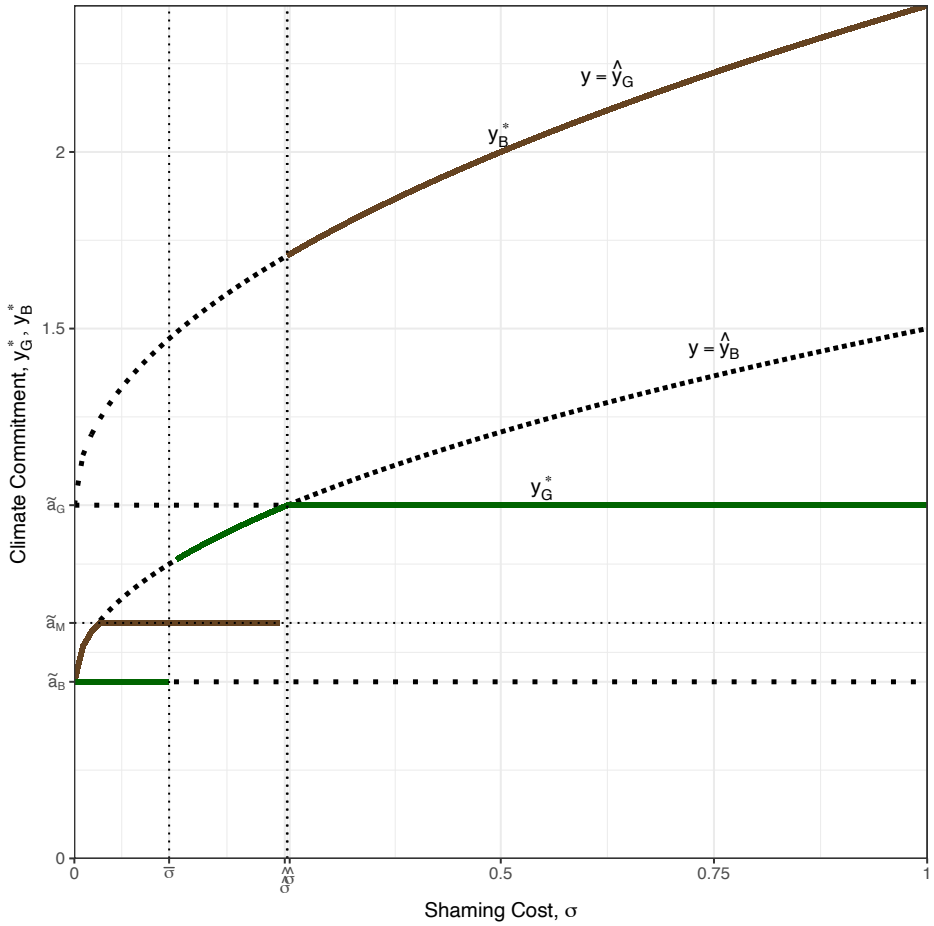
B's optimal commitment is

$$y_B^* = \begin{cases} \min\{\tilde{a}_M, \quad \hat{y}_B = \frac{1+\sqrt{2\sigma_B \lambda_B}}{\lambda_B}\} & \text{if } \sigma_G < \hat{\hat{\sigma}} \\ \hat{y}_G = \frac{1+\sqrt{2\sigma_G \lambda_G}}{\lambda_G} & \text{if } \sigma_G > \hat{\hat{\sigma}}. \end{cases}$$

Figure 4 plots the optimal climate commitments given the size of the shaming costs σ_g, as characterized in Proposition 3. First suppose that shaming costs are large ($\sigma_B \geq \hat{\sigma}$), as we considered in the analysis of Figure 3. By setting the commitment y_G^* at its ideal point, Green can remove any electoral bias in favor of the Brown party and simultaneously commit itself and the Brown party to its ideal point. This is clearly the first best scenario for the Green party. This is illustrated as the flat green line on the right-hand side of Figure 4. However, it requires that B's shaming cost σ_B is large enough that B would follow through on this commitment.

As B's shaming cost falls below the level sufficient to enforce G's ideal point ($\bar{\sigma} \leq \sigma_B \leq \hat{\sigma}$), then G's optimal commitment becomes less ambitious. G sets y_G^* to the highest downstream effort that commits B to compliance with the target, albeit this is less ambitious than G's ideal point. This is shown by the curved green segment on the line $y = \hat{y}_B$ in the figure. Such a commitment is partially beneficial for G in both policy and electoral terms: the target ties B's hands to implement a policy closer to G's ideal point and,

Figure 4: Climate Commitments and Shaming Costs for Office Seeking Parties



as seen in Figure 3, partially reduces B's electoral edge. If B's shaming costs are even smaller ($\sigma_B \leq \bar{\sigma}$), then from an electoral perspective G can do no better than set the commitment to B's ideal point. This is the flat green line on the left-hand side of the figure. When B's shaming cost is small then G can partially move the policy that B will implement above B's ideal point; however doing so makes B more electorally attractive to the median voter. When winning the election is a party's primary goal and B's shaming cost is small, G does not want to move B's downstream policy implementation above $\tilde{a}_B$. In summary, as shaming costs increase, the Green party can leverage the possibility of being shamed to enforce a more ambitious commitment (although never above its ideal point) while also reducing any electoral bias in B's favor.

Turning to B's optimal commitments, the Brown party has an electoral advantage because, *ex ante*, it would impose fewer costs to implement downstream climate policies on the median voter than the Green party would. Moreover, because the Green party is willing to commit to more ambitious pledges, the Brown

party can exploit this to their advantage by setting a lofty target they do not intend to implement themselves but the Green party would. Counterintuitively, this further B's electoral prospects. The intuition for this result is that, despite being costly for the median voter, the Green party would follow through on the commitment. As was shown in Figure 3, the electoral bias shifts in B's favor when commitments are large enough (between $\hat{y}_B$ and $\hat{y}_G$). This is the optimal strategy for the Brown party when G's shaming costs are sufficiently large ($\sigma_G > \hat{\tilde{\sigma}}$) and electoral incentives dominate. If elected, B would have to pay the shaming costs, but it improves B's odds at reelection.

Finally, when the Green party's shaming cost takes moderate values, the Brown party commits to credibly implement the median voter's ideal point. This is an electorally popular strategy because G would be implementing its ideal point, a downstream policy that imposes greater costs on the median voter. This is seen by the flat brown line toward the left-hand side of Figure 4. But if the shaming costs are insufficient to motivate the Brown party to implement the median voter's preferred policies, it picks a target to partially tie its hands, moving as close as it credibly can toward $\tilde{a}_M$, seen by the increasing brown line on the left-hand side of the figure. In summary, at low to moderate shaming costs, the Brown party will increase its commitment toward the preferences of the median voter. However, when shaming costs are sufficiently large, Brown has a dominant electoral strategy to commit to a pledge that it will paradoxically never implement but the Green party will; despite the downstream shaming this enhances Brown's electoral odds.

Propositions 2 and 3 demonstrate that commitments are always weakly increasing in the costs of being shamed σ_g. As internationally-imposed shaming costs for failing to meet national targets increase, leaders are willing to exert greater effort to meet such targets and therefore avoid being shamed. Note that this intuition does not rely on the fact that parties would be shamed at the same level. For example, if the Brown party paid scant attention to their international reputation, then the Green party can do little to tie the Brown party's hands. By contrast, the counterintuitive result that the Brown party can set a lofty climate target that it does not intend to meet and yet still increase its electoral odds is further strengthened if the Green party faces greater shaming costs. Indeed, the Brown party would choose targets of increasing ambition, which the Green party would comply with in order to avoid being shamed.

Discussion

We provide a model in which domestic political incentives shape international climate commitments. We demonstrate how domestic political competition can affect parties' willingness to commit to different pledges

and how the downstream implementation of policies to meet those pledges affects elections. Our model provides insights into the expected ambition of pledges, membership of the Paris Agreement, and the consequences of institutional strength.

Whether a nationally determined contribution is "ambitious" is often defined against some type of equity benchmark, e.g., if the NDC induces a nation to commit to reducing emissions commensurate with its "fair share" (Sælen et al. 2019). Robiou du Pont and Meinshausen (2018) propose five "equity approaches," and compare NDCs with fair share contributions to Paris's 1.5°C and 2°C reduction targets, finding that none of the world's top emitters submitted NDCs consistent with any of the equity approaches, and conclude that these NDCs are not ambitious enough. Rather than focus on fairness, our analysis implicitly proposes a positive measure of ambition, which is how much effort a government exerts above its ideal point given its climate commitment. We demonstrate that pledges can be ambitious because of their domestic political value. For example, when the cost of being shamed is relatively modest and leaders care primarily about holding office, the Brown party can propose a commitment that would force it to implement the median voter's ideal point. This commitment is ambitious because it is greater than the Brown party's ideal point, and it is electorally advantageous because it imposes fewer costs on the voters than the Green party's policy.

The Paris Agreement is often lauded because it attracted a wide membership, going against the conventional wisdom that international environmental agreements often garner only small coalitions (Calvo and Rubio 2013; Caparrós 2016; Harstad 2023*b*). This resembles the canonical "broader-deeper tradeoff" in the literature on international cooperation (Downs, Rocke and Barsoom 1998; Gilligan 2004; Johns 2014; Edry 2020). Broad agreements are sometimes castigated because they require little adjustment; however, our analysis suggests that leaders may be attracted to this type of agreement because of the way that climate commitments can affect policy outcomes. If leaders are primarily interested in the value of policy, they may be able to use commitments as a tool to tie the hands of their political rivals. Without *ex ante* commitments, leaders would simply implement their ideal points; however, the Green party can leverage the use of commitments within the Paris framework in order to bind the Brown party to more ambitious climate action.

Moreover, leaders also have incentives to ensure that Paris's enforcement mechanism, naming and shaming, has sufficient bite when reviewing climate pledges. While most extant literature lauds Paris for its flexibility and lack of formal enforcement (e.g., Bodansky 2016; Falkner 2016), the costs of being shamed need to be sufficiently high in order for leaders to exploit the agreement's structure for political gain. Indeed, if being shamed is costly, a policy-orientated Green party can force the Brown party into implementing the

Green party's ideal point through the choice of its climate commitment.

However, depending on leaders' incentives, the model also provides a cautionary tale of how the terms of strong international agreements can be exploited by domestic political actors and can facilitate outcomes counterproductive to international cooperative goals. When the shaming cost is high, an office-seeking Brown party, looking to enhance its electoral prospects, optimally commits to a lofty target, knowing full well that, if elected, it will not satisfy that pledge and will be shamed. However, if the Green party were to come to power after the election, they would pursue mitigation policies that would satisfy this target, imposing large costs on the voters in the process. Knowing this, voters are more likely to elect the Brown party in order to avoid paying the costs of intense mitigation measures. Leaders of anti-environmental parties can enhance their electoral prospects by promising something they cannot deliver.

Conclusion

This study probes the domestic political incentives that leaders have to choose climate commitments that affect the nature of future policymaking. Our formal model demonstrates the complexity of strategic calculations that leaders face when forging pledges, but also distills decisionmaking along two primary mechanisms: making commitments for policy value and making commitments for electoral gain. When policy concerns dominate, climate pledges can be valuable by tying the hands of political competitors, ensuring an enhanced level of mitigation effort. By contrast, if office concerns are more influential, then commitments can be exploited by leaders based on the expected costliness of downstream mitigation efforts relative to the median voter's willingness to pay. Paradoxically, this leads to lofty commitments made by leaders who never intend to fulfill them, but make such pledges in order to make environmentally-friendly parties electorally unattractive.

References

Aklin, Michaël and Matto Mildenberger. 2020. "Prisoners of the Wrong Dilemma: Why Distributive Conflict, Not Collective Action, Characterizes the Politics of Climate Change." *Global Environmental Politics* 20(4):4–27.

Aldy, Joseph Edgar, William Pizer, Massimo Tavoni, Lara Reis, Keigo Akimoto, Geoffrey Blanford, Carlo Carraro, Leon Clarke, James Edmonds, Gokul Iyer, Haewon McJeon, Richard Richels, Steven Rose and Fuminori Sano. 2016. "Economic tools to promote transparency and comparability in the Paris Agreement." *Nature Climate Change* .

Andrews, Talbot M and Scott E Bokemper. 2024. "The Road to Reelection is Paved with Good Intentions: Experiments on the Role of Outcomes and Intentions in Voting Behavior." *The Journal of Politics* .

Ansolabehere, Stephen and David M. Konisky. 2014. *Cheap and Clean: How Americans Think about Energy in the Age of Global Warming*. Cambridge: MIT Press.

Bagashka, Tanya and Randall W. Stone. 2013. "Risky Signals: The Political Costs of Exchange Rate Policy in Post-Communist Countries." *International Studies Quarterly* 57(3):519–531.

Battaglini, Marco and Bård Harstad. 2020. "The Political Economy of Weak Treaties." *Journal of Political Economy* 128(2):544–590.

Bechtel, Michael M. and Kenneth F. Scheve. 2013. "Mass support for global climate agreements depends on institutional design." *Proceedings of the National Academy of Sciences* 110(34):13763–13768.

Bergquist, Parrish, Matto Mildenberger and Leah C Stokes. 2020. "Combining climate, economic, and social policy builds public support for climate action in the US." *Environmental Research Letters* 15(5):054019.

Bodansky, Daniel. 2016. "The Legal Character of the Paris Agreement." *Review of European, Comparative & International Environmental Law* 25(2):142–150.

Buisseret, Peter and Dan Bernhardt. 2018. "Reelection and Renegotiation: International Agreements in the Shadow of the Polls." *American Political Science Review* 112(4):1016–1035.

Calvo, Emilio and Santiago J. Rubio. 2013. "Dynamic Models of International Environmental Agreements: A Differential Game Approach." *International Review of Environmental and Resource Economics* 6(4):289–339.

Caparrós, Alejandro. 2016. "Bargaining and International Environmental Agreements." *Environmental & Resource Economics* 65(1):5–31.

Casler, Don, Richard Clark and Noah Zucker. 2023. "Do Pledges Bind? The Mass Politics of International Climate Targets.".

Chapman, Terrence L, Johannes Urpelainen and Scott Wolford. 2013. "International bargaining, endogenous domestic constraints, and democratic accountability." *Journal of Theoretical Politics* 25(2):260–283.

Creamer, Cosette D and Beth A Simmons. 2019. "Do Self-Reporting Regimes Matter? Evidence from the Convention Against Torture." *International Studies Quarterly* 63(4):1051–1064.

Dai, Xinyuan. 2002. "Information Systems in Treaty Regimes." *World Politics* 54(4):405–436.

Dai, Xinyuan. 2007. *International Institutions and National Policies*. Cambridge: Cambridge University Press.

Dannenberg, Astrid, Marcel Lumkowsky, Emily K. Carlton and David G. Victor. 2023. "Naming and shaming as a strategy for enforcing the Paris Agreement: The role of political institutions and public concern." *Proceedings of the National Academy of Sciences* 120(40):e2305075120.

Downs, George W., David M. Rocke and Peter N. Barsoom. 1998. "Managing the Evolution of Multilateralism." *International Organization* 52(2):397–419.

Edry, Jessica. 2020. "Shallow commitments may bite deep: domestic politics and flexibility in international cooperation." *International Interactions* 46(5):669–695.

Egan, Patrick J, David M Konisky and Megan Mullin. 2022. "Ascendant Public Opinion: The Rising Influence of Climate Change on Americans' Attitudes about the Environment." *Public Opinion Quarterly* 86(1):134–148.

Falkner, Robert. 2016. "The Paris Agreement and the new logic of international climate politics." *International Affairs* 92(5):1107–1125.

Fearon, James D. 1994. "Domestic Political Audiences and the Escalation of International Disputes." *The American Political Science Review* 88(3):577–592.

Gaikwad, Nikhar, Federica Genovese and Dustin Tingley. 2022. "Creating Climate Coalitions: Mass Preferences for Compensating Vulnerability in the World's Two Largest Democracies." *American Political Science Review* 116(4):1165–1183.

Gazmararian, Alexander F. and Dustin Tingley. 2023. *Uncertain Futures: How to Unlock the Climate Impasse*. New York: Cambridge University Press.

Gazmararian, Alexander F., Matto Mildenberger and Dustin Tingley. 2024. "Public Opinion Foundations of the Clean Energy Transition.".

Gilligan, Michael J. 2004. "Is There a Broader-Deeper Trade-off in International Multilateral Agreements?" *International Organization* 58(3):459–484.

Hafner-Burton, Emilie M. 2005. "Trading Human Rights: How Preferential Trade Agreements Influence Government Repression." *International Organization* 59(3):593–629.

Hafner-Burton, Emilie M. 2008. "Sticks and Stones: Naming and Shaming the Human Rights Enforcement Problem." *International Organization* 62(4):689–716.

Hafner-Burton, Emilie M. 2012. "International Regimes for Human Rights." *Annual Review of Political Science* 15(1):265–286.

Hafner-Burton, Emilie M., Brad L. LeVeck and David G. Victor. 2017. "No False Promises: How the Prospect of Non-Compliance Affects Elite Preferences for International Cooperation." *International Studies Quarterly* 61(1):136–149.

Harrison, Rodrigo and Roger Lagunoff. 2017. "Dynamic Mechanism Design for a Global Commons." *International Economic Review* 58(3):751–782.

Harstad, Bård. 2023*a*. "Pledge and review bargaining." *Journal of Economic Theory* 207(C).

Harstad, Bård. 2023*b*. "Pledge-and-Review Bargaining: from Kyoto to Paris." *The Economic Journal* 133(651):1181–1216.

Hermwille, Lukas and Nicolas Kreibich. 2018. Article 6 and the Global Stocktake. JIKO Policy Paper 05/2018 Wuppertal Institute for Climate, Environment and Energy Wuppertal: .

Holz, Ceecee, Sivan Kartha and Tom Athanasiou. 2018. "Fairly sharing 1.5: national fair shares of a 1.5 °C-compliant global mitigation effort." *International Environmental Agreements: Politics, Law and Economics* 18(1):117–134.

Iida, Keisuke. 1993. "When and How Do Domestic Constraints Matter? Two-Level Games with Uncertainty." *The Journal of Conflict Resolution* 37(3):403–426.

Iida, Keisuke. 1996. "Involuntary defection in two-level games." *Public Choice* 89(3):283–303.

Jacquet, Jennifer and Dale Jamieson. 2016. "Soft but significant power in the Paris Agreement." *Nature Climate Change* 6(7):643–646.

Johns, Leslie. 2014. "Depth versus rigidity in the design of international trade agreements." *Journal of Theoretical Politics* 26(3):468–495.

Kennard, Amanda and Keith E. Schnakenberg. 2023. "Comment: Global Climate Policy and Collective Action." *Global Environmental Politics* 23(1):133–144.

Keohane, Robert O. and Michael Oppenheimer. 2016. "Paris: Beyond the Climate Dead End through Pledge and Review?" *Politics and Governance* 4(3):142–151.

Kotchen, Matthew J., Kevin J. Boyle and Anthony A. Leiserowitz. 2013. "Willingness-to-pay and policy-instrument choice for climate-change policy in the United States." *Energy Policy* 55:617–625.

Köke, Sonja and Andreas Lange. 2017. "Negotiating environmental agreements under ratification constraints." *Journal of Environmental Economics and Management* 83(C):90–106.

Leinaweaver, Justin and Robert Thomson. 2021. "The Elusive Governance of Climate Change: Nationally Determined Contributions as Commitments and Negotiating Positions." *Global Environmental Politics* 21(2):73–98.

Linos, Katerina and Tom Pegram. 2016. "The Language of Compromise in International Agreements." *International Organization* 70(3):587–621.

McAllister, Jordan H. and Keith E. Schnakenberg. 2022. "Designing the Optimal International Climate Agreement with Variability in Commitments." *International Organization* 76(2):469–486.

McGillivray, Fiona and Alastair Smith. 2008. *Punishing the Prince: A Theory of Interstate Relations, Political Institutions, and Leader Change*. Princeton: Princeton University Press.

Melnick, Justin and Alastair Smith. 2023. "International Negotiations in the Shadow of Elections." *Journal of Conflict Resolution* 67(7-8):1452–1481.

Milkoreit, Manjana and Kate Haapala. 2017. Designing the Global Stocktake: A Global Governance Innovation. Technical report Center for Climate and Energy Solutions Arlington: .

Milner, Helen V. 1997. *Interests, Institutions, and Information*. Princeton: Princeton University Press.

Milner, Helen V. and B. Peter Rosendorff. 1997. "Democratic Politics and International Trade Negotiations: Elections and Divided Government As Constraints on Trade Liberalization." *Journal of Conflict Resolution* 41(1):117–146.

Mo, Jongryn. 1995. "Domestic Institutions and International Bargaining: The Role of Agent Veto in Two-Level Games." *The American Political Science Review* 89(4):914–924.

Nemet, Gregory F. and Evan Johnson. 2010. "Willingness to Pay for Climate Policy: A Review of Estimates.".

Porter, Robert H. 1983. "Optimal cartel trigger price strategies." *Journal of Economic Theory* 29(2):313–338.

Putnam, Robert D. 1988. "Diplomacy and Domestic Politics: The Logic of Two-Level Games." *International Organization* 42(3):427–460.

Raiser, Kilian, Başak Çalı and Christian Flachsland. 2022. "Understanding pledge and review: learning from analogies to the Paris Agreement review mechanisms." *Climate Policy* 22(6):711–727.

Robiou du Pont, Yann and Malte Meinshausen. 2018. "Warming assessment of the bottom-up Paris Agreement emissions pledges." *Nature Communications* 9(1):4810.

Rosendorff, B. Peter and Helen V. Milner. 2001. "The Optimal Design of International Trade Institutions: Uncertainty and Escape." *International Organization* 55(4):829–857.

Rowan, Sam S. 2019. "Pitfalls in comparing Paris pledges." *Climatic Change* 155(4):455–467.

Slechten, Aurélie. 2020. "Environmental Agreements under Asymmetric Information." *Journal of the Association of Environmental and Resource Economists* 7(3):455–481.

Spilker, Gabriele and Tobias Böhmelt. 2013. "The impact of preferential trade agreements on governmental repression revisited." *The Review of International Organizations* 8(3):343–361.

Sælen, Håkon, Vegard Tørstad, Ceecee Holz and Tobias Dan Nielsen. 2019. "Fairness conceptions and self-determined mitigation ambition under the Paris Agreement: Is there a relationship?" *Environmental Science & Policy* 101:245–254.

Terman, Rochelle. 2023. *The Geopolitics of Shaming: When Human Rights Pressure Works—and When It Backfires*. Princeton: Princeton University Press.

Terman, Rochelle and Erik Voeten. 2018. "The relational politics of shame: Evidence from the universal periodic review." *The Review of International Organizations* 13(1):1–23.

Tingley, Dustin and Michael Tomz. 2022. "The Effects of Naming and Shaming on Public Support for Compliance with International Agreements: An Experimental Analysis of the Paris Agreement." *International Organization* 76:445–468.

Tørstad, Vegard, Håkon Sælen and Live Standal Bøyum. 2020. "The domestic politics of international climate commitments: which factors explain cross-country variation in NDC ambition?" *Environmental Research Letters* 15(2):024021.

Tørstad, Vegard and Vegard Wiborg. 2023. "Commitment Ambiguity and Ambition in Climate Pledges.".

UNFCCC. 2015. "Paris Agreement.".

Victor, David G., Marcel Lumkowsky and Astrid Dannenberg. 2022. "Determining the credibility of commitments in international climate policy." *Nature Climate Change* 12(9):793–800.

Appendix

Uniqueness of Optimal Mitigation Efforts

As eluded to in the main text, there can be multiple local maxima that satisfy the FOC characterized by equation 1. The government's payoff, $u_g(a_g, A; \lambda_g)$, is composed of two single peaked functions. In terms of policy, $a_g - \frac{a_g^2 \lambda_g}{2}$ is a concave function with a maximum at $a_g = \frac{1}{\lambda_g}$ that contributes the terms $1 - \lambda_g a_g$ to the FOC. The probability of being shamed is decreasing in a_g, and contributes the $\sigma_g \sqrt{\beta} \phi \left(\sqrt{\beta}(y - a_g) \right)$ term to the FOC.

$$foc = \frac{du_g(a_g, A; \lambda_g)}{da_g} = 1 - \lambda_g a_g^* + \sigma_g \sqrt{\beta} \phi(\sqrt{\beta}(y - a_g^*))$$

and the second-order condition (SOC)

$$soc = \frac{d^2 u_g(a_g, A; \lambda_g)}{da_g^2} = -\lambda_g + \sigma_g \beta \sqrt{\beta}(y - a_g^*) \phi(\sqrt{\beta}(y - a_g^*)).$$

When signals are imprecise the government's payoff is globally concave and so there is only a single solution to $foc = 0$, as formally stated in the following lemma.

Lemma A.1 *If $\beta < \frac{\sqrt{2e\pi}\lambda_g}{\sigma_g}$ (or equivalently $\sigma_g < \frac{\sqrt{2e\pi}\lambda_g}{\beta}$), then the government's payoff $u_g(a_g, A; \lambda_g)$ is globally concave for any y, there is a unique solution to $foc = 0$ and $a_g^* \geq \frac{1}{\lambda_g}$.*

Proof of Lemma A.1: The second order condition is $soc = -\lambda_g + \sigma_g \beta \sqrt{\beta}(y - a_g) \phi(\sqrt{\beta}(y - a_g))$, which has a maximum of $\frac{\beta \sigma_g}{\sqrt{2e\pi}} - \lambda_g$ at $y - a_g = \frac{1}{\sqrt{\beta}}$. Hence if $\beta < \frac{\sqrt{2e\pi}\lambda_g}{\sigma_g}$ then soc is always negative and the government's optimization is globally concave and foc is decreasing in a_g. At $a_g = \frac{1}{\lambda_g}$, $foc \geq 0$ and as $a_g \to \infty$, $foc \to -\infty$, therefore there is a unique $a_g^* \geq \frac{1}{\lambda}$ such that $foc = 0$. ■

If signals are more precise then the government's utility function, $u_g(a_g, A; \lambda_g)$, is potentially two peaked with a peak around $a_g = \frac{1}{\lambda_g}$ and another peak around $a_g = y$. If y is relatively close to $\frac{1}{\lambda_g}$, then these two peaks coincide resulting in the aggregate $u_g(\cdot)$ being single peaked. In contrast if y is relatively large compared to $\frac{1}{\lambda_g}$, then $u_g(a_g, A; \lambda_g)$ is two peaked and there are two local maxima that satisfy the $foc = 0$ (and $soc < 0$). Further since $u_g(\cdot)$ is continuous, if there are two local maxima, then there must also be a local minimum between them that satisfies $foc = 0$ and $soc > 0$. The following lemma exploits this graphical exposition of the shape of $u_g(\cdot)$.

The first two conditions show that when y is relatively extreme (less than $\frac{1}{\lambda_g}$ or greater than $\frac{1+\sqrt{2\lambda_g \sigma_g}}{\lambda_g}$), then, with precise signals, the government's effort is close to $\frac{1}{\lambda_g}$. The third condition exploits the fact that

if there are two local maxima that satisfy $foc = 0$, then there must also be a local minimum between them. If signals are imprecise, then no such minimum can exist and therefore there is a unique local maximum. In contrast, if signal are precise, then two local maxima that satisfy $foc = 0$ can exist and therefore a_g^* can be discontinuous in y.

Lemma A.2 *1. If $y \leq \frac{1}{\lambda_g}$, then $a_g^* \geq \frac{1}{\lambda_g}$.*

2. If $y \geq \frac{1+\sqrt{2\lambda_g\sigma_g}}{\lambda_g}$, then $a_g^* \in [1/\lambda_g, y)$.

3. $\beta < \frac{4\lambda_g^2}{(\lambda_g y-1)^2}$ is sufficient to ensure there is a unique local maximum that satisfies $foc = 0$ and a_g^* is continuous in y. If $\beta > \frac{4\lambda_g^2}{(\lambda_g y-1)^2}$ then there can be two maxima that satisfy $foc = 0$ and a_g^* can be discontinuous in y.

4. As $\beta \to \infty$, $a_g^* \to max\{y, \frac{1}{\lambda_g}\}$ if $y < \frac{\sqrt{2}\sqrt{\lambda_g}\sqrt{\sigma_g}+1}{\lambda_g}$; and $a_g^* \to \frac{1}{\lambda_g}$ if $y > \frac{\sqrt{2}\sqrt{\lambda_g}\sqrt{\sigma_g}+1}{\lambda_g}$.

Proof of Lemma A.2: For part 1, if $y \leq \frac{1}{\lambda_g}$, then for $a_g < 1/\lambda_g$ the government's payoff is strictly increasing in a_g. For $a_g = 1/\lambda_g$, the $foc \geq 0$ and for all $a_g > 1/\lambda_g$, $soc < 0$, so foc is strictly decreasing in a_g for all $a_g \geq 1/\lambda_g$ and therefore the foc can only cross zero once.

For part 2 consider the following limiting cases. The government can always play $a_g = \frac{1}{\lambda_g}$ and get a payoff at least as big as $\frac{1}{2\lambda_g} - \sigma_g$. In contrast suppose the government plays $a_g \geq y$ and take the limiting case that playing $a_g = y$ fully avoids shame (limiting case as $\beta \to \infty$). The payoff from this effort is less than or equal to $y - \frac{y^2\lambda_g}{2}$. Comparing these payoffs, the former is larger if $y \geq \frac{1+\sqrt{2\lambda_g\sigma_g}}{\lambda_g}$. Hence when this condition holds, the government prefers to play some $a_g \in [1/\lambda_g, y)$, than any $a_g \geq y$.

For part 3, when $foc = 0$ holds then, $\lambda_g a_g^* - 1 = \sigma_g\sqrt{\beta}\phi(\sqrt{\beta}(y - a_g^*))$. Substitute the RHS into SOC: $soc = -\lambda_g + \beta(y - a_g^*)(\lambda_g a_g^* - 1)$. Since the $u_g(\cdot)$ is continuous in a_g, there can only be two local maxima if there is also a local minimum between them. The soc expression is maximized by $a_g = \frac{y+1/\lambda_g}{2}$ which yields a maximum of $-\lambda_g + \frac{\beta+\beta\lambda_g^2y^2-2\beta\lambda_g y}{4\lambda_g}$. Hence provided that $\beta < \frac{4\lambda_g^2}{(\lambda_g y-1)^2}$, the soc expression is negative for all $foc = 0$ and so there cannot be a local minimum. Absent a local min there must be a unique maximum. In contrast if signals are relatively precise, $\beta > \frac{4\lambda_g^2}{(\lambda_g y-1)^2}$, then there can be two local maxima that satisfy $foc = 0$ and the best effort a_g^* can be discontinuous in y.

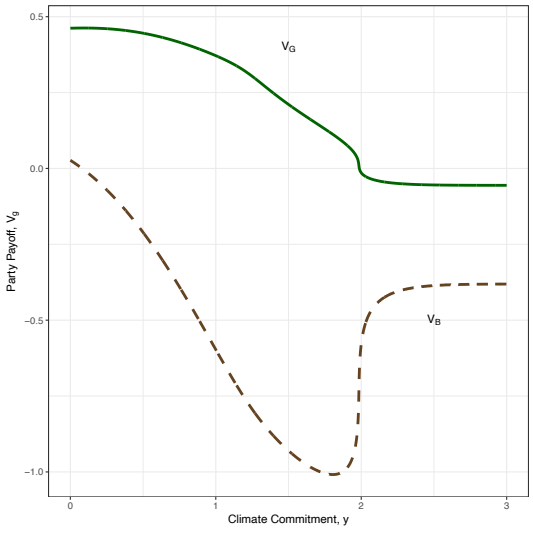
Part 4 is simply the limiting case elaborated on in the text. If $a_g < y$, then $u_g(a_g) = A - \frac{\lambda_g a_g^2}{2} - \sigma_g$ which is maximized by $a_g = \frac{1}{\lambda_g}$. If $a_g > y$, then $u_g(a_g) = A - \frac{\lambda_g y^2}{2}$, which for $y > \frac{1}{\lambda_g}$ is maximized by $a_g = y$. The condition $y = \frac{\sqrt{2}\sqrt{\lambda_g}\sqrt{\sigma_g}+1}{\lambda_g}$ follows directly from equating these payoffs. ■

Party Payoffs in the General Model

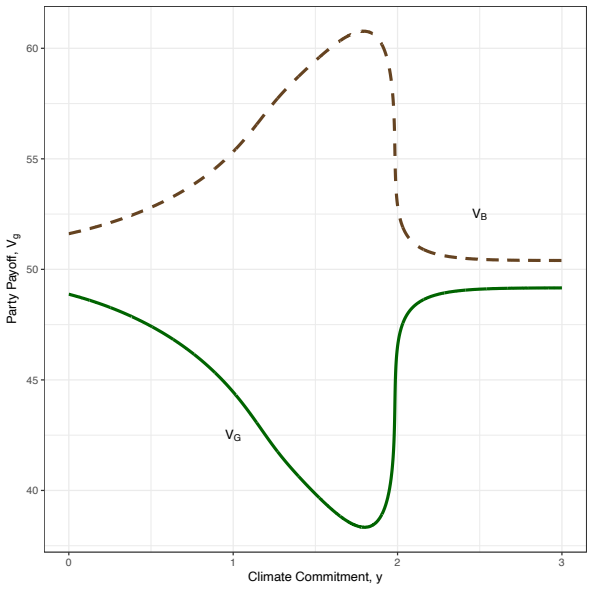
We plot parties' payoffs V_B and V_G in Figure A.1. The left panel of the figure shows party payoffs as a function of y if parties only care about policy outcomes. By setting y, a party can influence the policy choice of the other party. For instance, G can tie B's hands in terms of enacting greater mitigation efforts after the election. As y becomes more ambitious, B's payoff decreases quite substantially as it exerts effort further and further from its ideal point to meet the pledge. By contrast, since G would be willing to implement more ambitious mitigation strategies *ex ante*, its payoff decreases less dramatically as it incurs the costs of exerting effort to meet an increasingly ambitious commitment. For sufficiently high y, it becomes too costly for either party to meet the commitment, and they revert to implementing their ideal points, knowing that it is likely that they will be shamed. In this case, parties generically prefer a lower commitment so it will be easy for them to both implement their ideal point and avoid shaming.

Figure A.1: Party Payoffs as a Function of Commitments

(a) Party Payoffs: Policy Incentives



(b) Party Payoffs: Office Incentives



In the right panel of Figure A.1, we plot party payoffs if their main incentive in pursuing climate commitments is to remain in office. Parties' considerations change dramatically when they select commitments in order to maximize electoral success. As we describe in the limiting cases, despite their *ex ante* distaste for climate action, the Brown party may have incentives to set a climate commitment that is highly ambitious. In so doing, B can set a target that is too high for them to meet, knowing they will likely be shamed if they win the election, but G will attempt to pursue it. G's adventurous mitigation efforts then

appear extremely costly for the voter, who knows that B, in failing to meet the commitment, will exert effort closer to the voter's ideal point. Office-holding concerns can therefore generate counterintuitive cases in which anti-climate governments set more ambitious climate commitments than pro-climate governments, knowing full well that they will not be honored, but are made in order to leverage the fact that pro-climate governments would become less electorally attractive to voters.

Generalization and Proofs of Winning Office Limiting Case

For the precise shaming limiting case ($\beta \to \infty$), it is useful to restate some definitions and define several new quantities:

1. $\hat{y}_g = \frac{1+\sqrt{2\sigma_g \lambda_g}}{\lambda_g}$ is the highest commitment g will implement before preferring to implement its ideal point and be shamed.

2. $\hat{\sigma} = \frac{(\lambda_B - \lambda_G)^2}{2\lambda_G^2 \lambda_B}$ is the minimum shaming cost such that B prefers to implement G's ideal point rather than implement its own ideal point and be shamed: $u_B(\tilde{a}_G) = u_B(\tilde{a}_B) - \hat{\sigma}$.

3. $\bar{y} = \frac{2\lambda_B - \lambda_M}{\lambda_B \lambda_M}$ is the policy commitment (above $\tilde{a}_B$) such that, if implemented, the median voter would be indifferent between $\bar{y}$ and B's ideal point.

4. $\bar{\sigma} = \frac{2(\lambda_M - \lambda_B)^2}{\lambda_B \lambda_M^2}$ is the smallest shaming cost such that B would implement $\bar{y}$ if elected (i.e. $u_B(\bar{y}) = u_B(\tilde{a}_B) - \bar{\sigma}$).

5. $\bar{\bar{\sigma}} = \frac{(\lambda_M - \lambda_B)^2}{2\lambda_M^2 \lambda_B}$ is that smallest shaming cost such that B can implement the median voter's ideal point: $u_B(\tilde{a}_M) = u_B(\tilde{a}_B) - \bar{\bar{\sigma}}$.

6. $\hat{\hat{\sigma}}$ is defined such that $\Delta(\tilde{a}_G, \tilde{a}_M; y = \tilde{a}_M) = \Delta(\hat{y}_G, \tilde{a}_B; y = \hat{y}_G)$. This is the smallest shaming cost such that the largest commitment that G can credibly implement produces the same electoral bias as B committing to the median voter's ideal point ($y = \tilde{a}_M$).

To limit the analysis to substantively interesting cases, we make the following assumption:

Assumption 1 *If* $\sigma_B > \hat{\sigma}$ *then* $\sigma_G > \frac{(\lambda_B - \lambda_G)^2 - 2\sqrt{2}\sqrt{\lambda_B}\lambda_G(\lambda_B - \lambda_G)\sqrt{\sigma_B}}{2\lambda_B^2 \lambda_G} + \frac{\lambda_G \sigma_B}{\lambda_B}$.

This condition ensures that when B can commit to a policy above $\tilde{a}_G$ that $\hat{y}_G > \hat{y}_B$, which substantively means that the Green party can implement larger commitments than the Brown party. The condition is only violated if Brown's shaming cost vastly exceeds Green's, such that Brown can commit to providing more policy than Green. Such a case seems substantively unlikely.

The proposition below specifies the optimal commitments for office-seeking parties.

Proposition A.1 *Let* $\beta \to \infty$, $\Psi \to \infty$, $F' > 0$ *and Assumption 1 holds. G's optimal climate commitment is*

$$y_G^* = \begin{cases} y \leq \tilde{a}_B \text{ or } y \in (\hat{y}_B, \tilde{a}_G] & \text{if } \sigma_B < \bar{\sigma} \\ \hat{y}_B = \frac{1+\sqrt{2\sigma_B \lambda_B}}{\lambda_B} & \text{if } \bar{\sigma} \leq \sigma_B \leq \hat{\sigma} \\ \tilde{a}_G & \text{if } \sigma_B > \hat{\sigma}. \end{cases}$$

If $\sigma_B \geq \bar{\bar{\sigma}}$ then B's optimal commitment is

$$y_B^* = \begin{cases} \tilde{a}_M & \text{if } \sigma_G \leq \hat{\hat{\sigma}} \\ \hat{y}_G = \frac{1+\sqrt{2\sigma_G\lambda_G}}{\lambda_G} & \text{if } \sigma_G > \hat{\hat{\sigma}}. \end{cases}$$

If $\sigma_B < \bar{\bar{\sigma}}$ then B's optimal commitment is

$$y_B^* = \begin{cases} \hat{y}_B = \frac{1+\sqrt{2\sigma_B\lambda_B}}{\lambda_B} & \text{if } \Delta(\tilde{a}_G, \hat{y}_B; \hat{y}_B) \geq \Delta(\hat{y}_G, \tilde{a}_B; \hat{y}_G) \\ \hat{y}_G = \frac{1+\sqrt{2\sigma_G\lambda_G}}{\lambda_G} & \text{if } \Delta(\tilde{a}_G, \hat{y}_B; \hat{y}_B) < \Delta(\hat{y}_G, \tilde{a}_B; \hat{y}_G) \end{cases}$$

Proof of Proposition A.1: Since officeholding dominates, Green seeks to maximize $\Delta(a_G^*, a_B^*; y)$; while Brown seeks to minimize $\Delta(a_G^*, a_B^*; y)$. We consider each case.

First suppose that G is the incumbent and B's shaming cost is small: $\sigma_B < \bar{\sigma}$. All the equilibrium commitments ($y \leq \tilde{a}_B$ or $y \in (\hat{y}_B, \tilde{a}_G]$) result in B and G each implementing their ideal point. Can G do better? No, if G's commitment is above $\tilde{a}_G$ then either G implements a policy above its ideal point (which is bad both in terms of policy and electability) or G is shamed. So G never profits by $y > \tilde{a}_G$. If G's commitment is $y \in (\tilde{a}_B, \hat{y}_B]$ then if elected B would implement this commitment, which is closer to the median voter's ideal point than $\tilde{a}_B$; this would reduce the electoral bias and harm G's electoral prospects.

Second, consider the case of a moderate shaming cost: $\bar{\sigma} \leq \sigma_B \leq \hat{\sigma}$. The largest commitment that B would implement is $\hat{y}_B = \frac{1+\sqrt{2\sigma_B\lambda_B}}{\lambda_B}$, which is above the median voter's ideal point and above $\bar{y}$. In this range, the electoral bias is increasing in y, subject to y being implemented by B. Hence G maximizes electoral bias by a commitment to the maximizes the policy that B implements.

Finally if $\sigma_B > \hat{\sigma}$, then any commitment $y \in [\tilde{a}_G, \hat{y}_B]$ results in both parties implementing the same post election policy, which maximizes the electoral bias. Within this set of electorally optimal policies, G prefers that its ideal point is implemented. Hence $y_G^* = \tilde{a}_G$.

Now consider B's optimal commitments. The analysis is split into two cases. First suppose that B's shaming cost is sufficiently large that B can implement the median voter's ideal point: $\sigma_B \geq \bar{\bar{\sigma}}$. As we saw from the discussion of Figure 3, for all $y \leq \tilde{a}_G$, the median voter's ideal point minimizes $\Delta(a_G^*, a_B^*; y)$. If B proposes $y > \tilde{a}_G$, then $\Delta(a_G^*, a_B^*; y)$ is minimized by pledging $\hat{y}_G$, the largest policy that G will implement. Note that by assumption 1, at this pledge, B would renege and be shamed. Thus, B's optimal choice will be a policy that minimizes one of the two following electoral biases, $\Delta(\hat{y}_G, \tilde{a}_B^*; \hat{y}_G)$ or $\Delta(\tilde{a}_G, \tilde{a}_M; y = \tilde{a}_M)$. When G's shaming cost is large ($\sigma_G > \hat{\hat{\sigma}}$), then the former is the optimal as it produces the greatest electoral

bias in B's favor; and when shaming cost is smaller then the latter is optimal.

Second, suppose B's shaming cost is insufficient for B to implement the median voter's ideal point: $\sigma_B < \bar{\bar{\sigma}}$. The analysis is similar to that case above, however, now B cannot commit to the median voter's ideal policy. Instead B picks between the largest policy that it can implement ($\hat{y}_B$) or the largest policy that G can implement. The electoral biases for these pledges are $\Delta(\tilde{a}_G, \hat{y}_B; \hat{y}_B)$ and $\Delta(\hat{y}_G, \tilde{a}_B; \hat{y}_G)$, respectively. Given the primacy of office holding, B selects the pledge with the largest electoral bias in B's favor. ■

In Proposition 3, G has a range of optimal commitments when $\sigma_B < \bar{\sigma}$; however, all such commitments result in an observationally equivalent outcome where the commitment does not affect B's downstream effort to implement policy at its ideal point. To plot Figure 4, we use the equilibrium refinement that selects the largest commitment that G would implement (that results in no shaming for B).