Shaming Paris: A Political Economy of Climate Commitments

Justin Melnick and Alastair Smith

March 6, 2024

Abstract

This paper explores leaders' incentives to set climate commitments and subsequently exert costly effort toward implementing those commitments with a formal model. Governments first make commitments to reduce emissions and then implement policies to meet their commitments. If a noisy signal of the government's mitigation efforts is less than the government's pledge then the incumbent party is liable to be "shamed." We characterize how the threat of shaming increases efforts to meet climate commitments and how domestic politics shapes both efforts and initial commitments. Two parties, Green and Brown, compete electorally; the Green party faces a lower cost for emissions reductions than the Brown party. Voters factor in the equilibrium policies that each party will implement given the pledge initially made. When pledging at Paris, parties consider how their commitments can tie the hands of each party and how the commitments affect who will win the election.

Word Count: 7814 (text) + 952 (references) + 878 (appendix) = 9644 (total)

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, up from 112 in 2019, with approximately 90% of these pledges including measures to raise their nation's climate mitigation capabilities (UNDP 2021). 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 Prime Minister Rishi Sunak, also a member of the Conservative Party, 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 anticipated costs from the green transition away from voters in the run-up to the country's next election: while the Labour Party claimed it would not restore the original deadline for phasing out gas boilers, it would revert back to the initial policy on the sale of petrol and diesel cars.[5] Sunak has hoped to pit the Conservatives as a cheaper, albeit less green, alternative to consumers than a potential Labour government. Both parties seem to be striking a balance between meeting the UK's climate commitment without imposing substantial costs on voters.

We present a formal model that probes the electoral motivations behind the setting of climate targets. In

[1] https://unfccc.int/sites/default/files/NDC/2022-09/UK%20NDC%20ICTU%202022.pdf
[2] https://www.gov.uk/government/news/uk-enshrines-new-target-in-law-to-slash-emissions-by-78-by-2035
[3] https://www.gov.uk/government/news/uk-becomes-first-major-economy-to-pass-net-zero-emissions-law
[4] https://www.theguardian.com/environment/2023/sep/20/rishi-sunak-confirms-rollback-of-key-green-targets
[5] https://www.theguardian.com/politics/2023/sep/20/sunak-u-turn-on-green-policies-puts-labour-in-difficult-position

the model, 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. 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 determine whether to support the Green or the Brown party given the downstream mitigation measures induced by the nation's pledge. At the end of the model, nations observe noisy signals of each leader's effort into mitigation 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). This institutional arrangement makes understanding the domestic political considerations surrounding the setting of commitments paramount. Second is the idea that leaders set pledges, their nationally determined contributions, 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 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 some type of reputational cost from being "named and shamed" for failing to comply with their targets (Bodansky 2016; Jacquet and Jamieson 2016). Along these lines, we consider how the domestic electoral competition between political parties entices leaders to set different types of commitments in the shadow of possible shaming.

Our analysis uncovers two relevant mechanisms through which domestic politics affect climate commitments. First, 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 anticipated 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. This is because, if Paris's institutional effects are strong, meaning the costs from shaming are sufficiently large, the Green party is willing to fulfill more ambitious pledges while the Brown party is 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. 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.

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 2023). 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 predicting 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 and implement international treaties (Buisseret and Bernhardt 2018; Melnick and Smith 2023). Similarly, 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.

More specifically, our model provides a domestically-microfounded story of the implementation of the Paris Agreement. Harstad (2023*b*) presents a dynamic bargaining model that varies the weight that countries place on their own contribution versus other nations' contributions, documenting the conditions under which the former scenario, corresponding to the Paris Agreement, yields more ambitious commitments than the latter, representing the Kyoto Protocol. Like Harstad's model, we endogenize the choice of initial commitment as in the Paris framework. However, we expand upon this by considering the political incentives that leaders may have to make such commitments. Other models of international climate cooperation seek to capture how Paris's role in disseminating information affects the scope for ambitious contributions (Harrison and Lagunoff 2017; Slechten 2020; McAllister and Schnakenberg 2022; Melnick 2024).

We also identify the political forces that affect the implementation of international agreements like the Paris Climate Accords that rely on self-reported commitments and "naming and shaming" as an enforcement mechanism. Extant literature has argued that the effectiveness of international regimes based on self-reporting processes requires sufficient transfer of information to experts and peers (Creamer and Simmons 2019; Raiser, Çalı and Flachsland 2022). Empirical studies of mass publics (Tingley and Tomz 2022; Casler, Clark and Zucker 2023) and policy elites (Dannenberg et al. 2023) demonstrate that individuals view naming and shaming as an adequate means of enforcing international climate commitments. More generally, Hafner-Burton, LeVeck and Victor (2017) demonstrate that policy elites prefer to join international agreements with informal enforcement mechanisms.

Finally, we complement a burgeoning empirical literature on the effects of the Paris Agreement and the determinants of nationally determined contributions. Tørstad and Wiborg (2023) find a negative correlation between the ambiguity of NDCs and their ambition. They also 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

Our primary empirical referent is the Paris Agreement, the current framework for international climate governance. Paris 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*). 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 (UNFCCC 2015). 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 then choose whether to abide by their targets in the shadow of informal enforcement often referred to as "naming and shaming."

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 they are judged to have failed to fulfill their commitment. This cost is larger if leaders anticipate greater reputational sanction for breaching their commitment, and larger shaming costs can entice leaders to fulfill larger commitments. However, in a world with imperfect monitoring (Porter 1983; Green and Porter 1984), 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: nations observe 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 $\sigma \in \mathbb{R}_+$. 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.

"Naming and shaming" is the cornerstone enforcement mechanism of the Paris Agreement (Falkner 2016). We are agnostic as to the form with which these shaming costs materialize: they may originate from international measures like increased tariffs (Nordhaus 2015) or from domestic disapproval or threat of electoral sanction (Tingley and Tomz 2022; Casler, Clark and Zucker 2023; Melnick 2024). What is important is that, as observed by Dannenberg et al. (2023), policy elites believe these costs exist.

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,

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

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. If the government wins the election, it enjoys the benefit $\Psi > 0$. Notice also that only the party in power incurs the shaming cost σ 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 solve for the model's 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 then 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

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

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

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

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

and the second-order condition (SOC)

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

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 2 and 3).

The FOC illustrates that parties weigh the marginal costs of exerting effort with their global marginal benefits and the possibility of being shamed. It is clear that parties would never exert effort less than their ideal point; Paris hopes to induce governments to exert effort greater than their ideal point, or $a_g^* > \tilde{a}_g$. Whether the institution has any bite will depend on the distance between the commitment y and the leader's ideal point $\tilde{a}_g$, the costs of being shamed σ, and the precision of the signal β.

In general, parties 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 minimize the probability of being shamed.

The comparative static of the commitment y on optimal effort a_g^* puts a finer point on the idea that the effect of commitments on effort is non-monotonic:

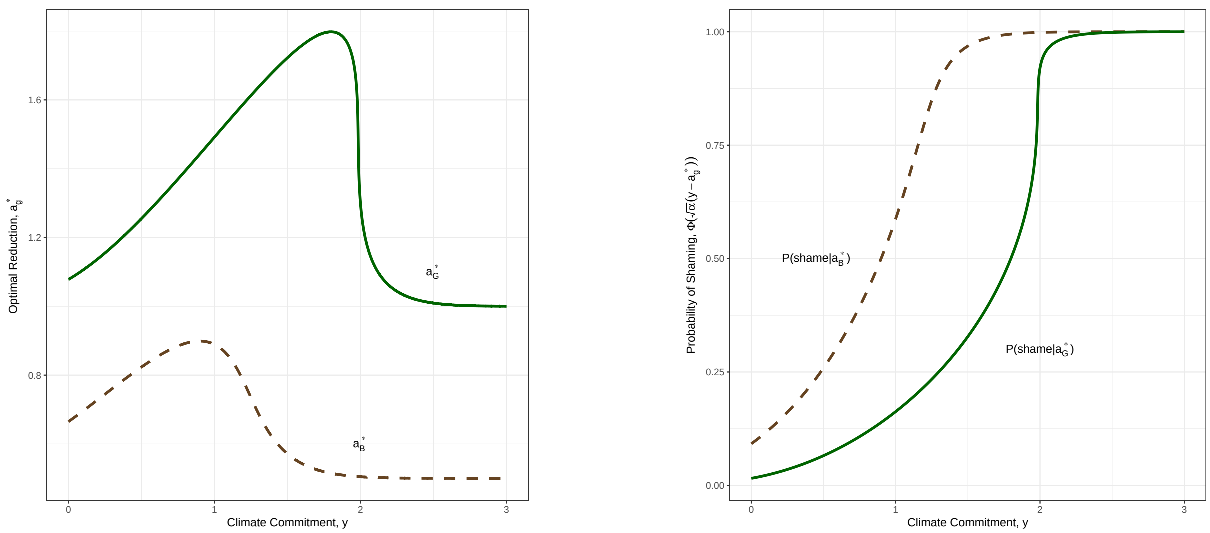
$$\frac{da_g^*}{dy} = \frac{-\sigma\beta\sqrt{\beta}(y - a_g^*)\phi(\sqrt{\beta}(y - a_g^*))}{\lambda_g - \sigma\beta\sqrt{\beta}(y - a_g^*)\phi(\sqrt{\beta}(y - a_g^*))}.$$

The sign of this expression is determined by the relationship between y and a_g^*. If $a_g^* < y$ then a_g^* is decreasing in y and if $a_g^* > y$ then a_g^* is increasing in y.

Figure 1 plots the optimal efforts of each party and the likelihood of being shamed as a function of the climate commitment y.[6] The left panel plots G's optimal effort in green (solid line) and B's optimal effort in brown (dashed line). As mentioned above, parties' optimal efforts are non-monotonic in the commitment y. If y is not too large then governments may find it in their interest to comply with the target as a means of avoiding shame. However, this incentive dissipates if the target is set too ambitiously as the cost of the effort to avoid being shamed is too high; parties resign themselves to being shamed and revert to implementing an effort strategy 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: $\frac{da_g^*}{d\lambda_g} = -\frac{a_g^*}{\lambda_g - \sigma\beta\sqrt{\beta}(y-a_g^*)\phi(\sqrt{\beta}(y-a_g^*))} < 0$. Furthermore, as illustrated in the right panel of the figure, exerting greater effort means that G will be shamed with a smaller probability than B.

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

(a) Optimal Efforts (b) Probability of Shaming



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 concerns affect the electoral outcome.

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

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

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

Substituting for the voter's utility function we can state

$$\Delta = \Delta(a_G^*, a_B^*; y) = a_G^* - a_B^* - \lambda_M \left(a_G^{*^2} - a_B^{*^2} \right) = (1 - (a_G^* + a_B^*)\lambda_M)(a_G^* - a_B^*).$$

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. The voter then adjudicates the relative costliness of G and B's expected policies against other electorally salient issues. 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 $F(\Delta)$.

We plot this bias Δ (dotted line) in Figure 2 along with parties' equilibrium efforts as a function of possible climate commitments y. Clearly, 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. It is also useful to know that, 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, $\Delta(\tilde{a}_G, \tilde{a}_B) = -\frac{(\lambda_B - \lambda_G)^3}{4\lambda_G^2 \lambda_B^2} < 0$. Formally, this 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. Substantively, this bias may represent inertia on behalf of the voters in their willingness to pay for more ambitious climate action relative to a status quo. Given empirical evidence suggesting electoral backlash to implementing costly mitigation strategies (e.g., Stokes 2016), we find this bias in favor of the Brown party to be substantively commensurable.

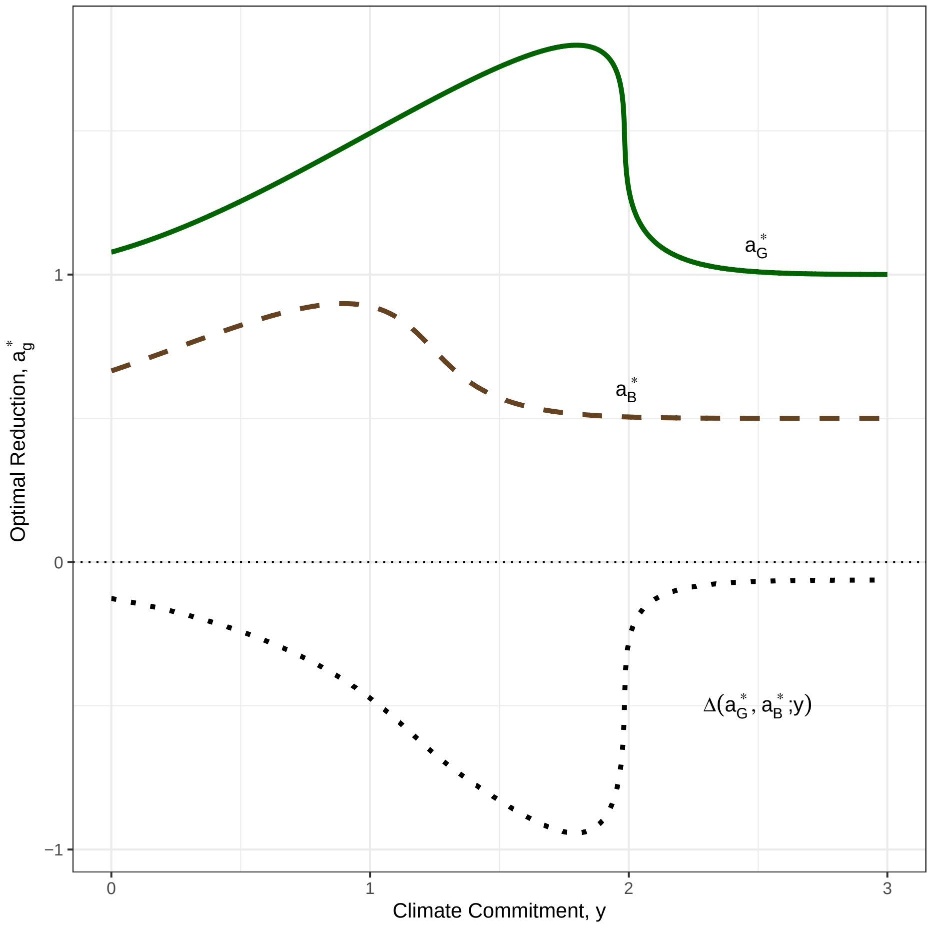


Figure 2: Electoral Bias and Efforts as a Function of Commitments

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:

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

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

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

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

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

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 $F(\Delta(a_G^*, a_B^*; y))$.*

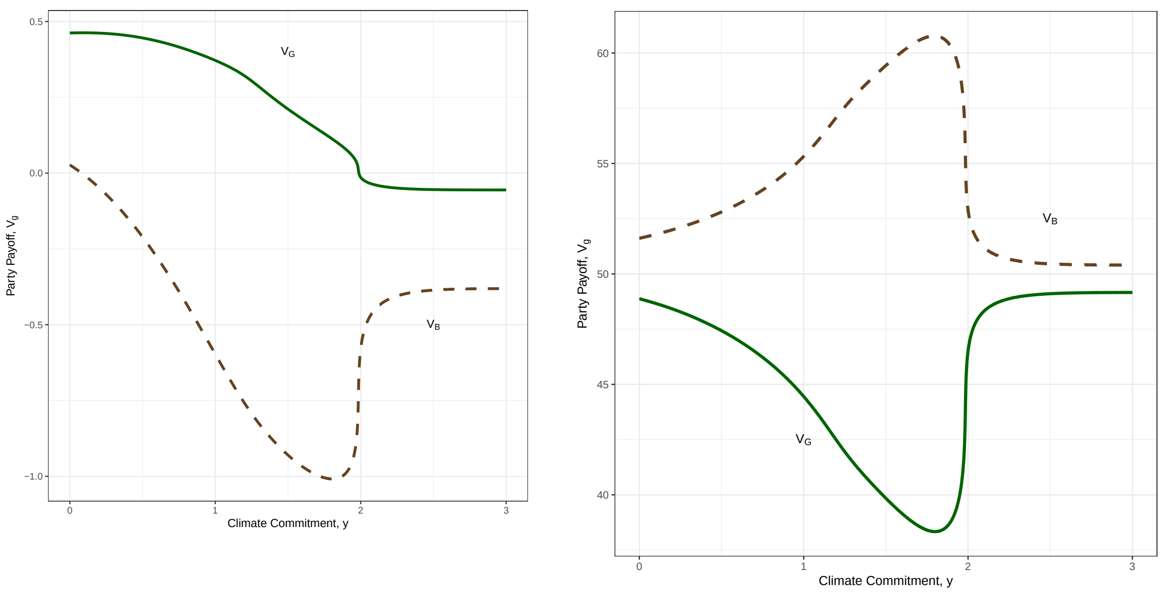
A party's choice of the commitment y will affect several factors that are driven by how pledges affect downstream mitigation efforts after the election. Climate commitments can thus be useful in policy terms, as parties may be able to tie the hands of their competitors through their choice of pledge. 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.

We plot parties' payoffs V_B and V_G in Figure 3. 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 low commitment so it will be easy for them to both implement their ideal point and avoid shaming.

Figure 3: Party Payoffs as a Function of Commitments

(a) Party Payoffs: Policy Incentives

(b) Party Payoffs: Office Incentives



In the right panel of Figure 3, 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 below, 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.

As Figure 3 makes clear, it is difficult to isolate the substantive impact of leaders' competing policy and office incentives in this general setting. By looking at a series of the limiting cases we can effectively isolate the influence of each factor on climate commitments.

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. Leaders' optimal efforts are fairly simple in this case: the winner of the election will either comply with the target or will implement their ideal effort level. If the target is low, then leaders can implement their ideal point and avoid shaming. Increasing the ambition of the climate commitment means leaders face a trade-off between implementing the target and rebuking it, implementing their ideal point instead and incurring the costs of being shamed. Leaders prefer to comply with the target instead of implementing their ideal point and getting shamed whenever $y \leq \hat{y}_g = \frac{1+\sqrt{2\lambda_g \sigma}}{\lambda_g}$. The definition of $\hat{y}_g$ follows from the largest pledge that g would implement rather than implement their ideal point and be shamed: $u_g(\hat{y}_g) = u_g(\frac{1}{\lambda_g}) - \sigma$.

For intermediate pledges, $\tilde{a}_g \leq y \leq \hat{y}_g$, it is optimal to exert effort in line with the target. But 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*

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

Tying Hands

Suppose first that parties choose climate commitments solely for their policy value. In particular, 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 elections are insensitive to climate commitments, the probability of G winning the election is effectively a constant $p \in [0, 1]$ such that G's payoff is

$$V_G(y) = p(u_G(a_G^*) - \mathbb{1}_{y > a_G^*}\sigma) + (1-p)u_G(a_B^*),$$

with B's payoff defined analogously. 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 exert effort closer to G's ideal point. G never sets a target above its own ideal point, because it would always prefer a policy $y = \tilde{a}_G$. 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 $\sigma \geq \hat{\sigma}$, G can 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 ($\sigma < \hat{\sigma}$), G cannot induce B to exert effort at G's ideal point $\tilde{a}_G$; B would rather be shamed. Instead, G ties B's hands to the greatest extent possible by setting $y_G^* = \hat{y}_B = \frac{1+\sqrt{2\lambda_B \sigma}}{\lambda_B}$. 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.

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*

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

B's optimal commitment is any

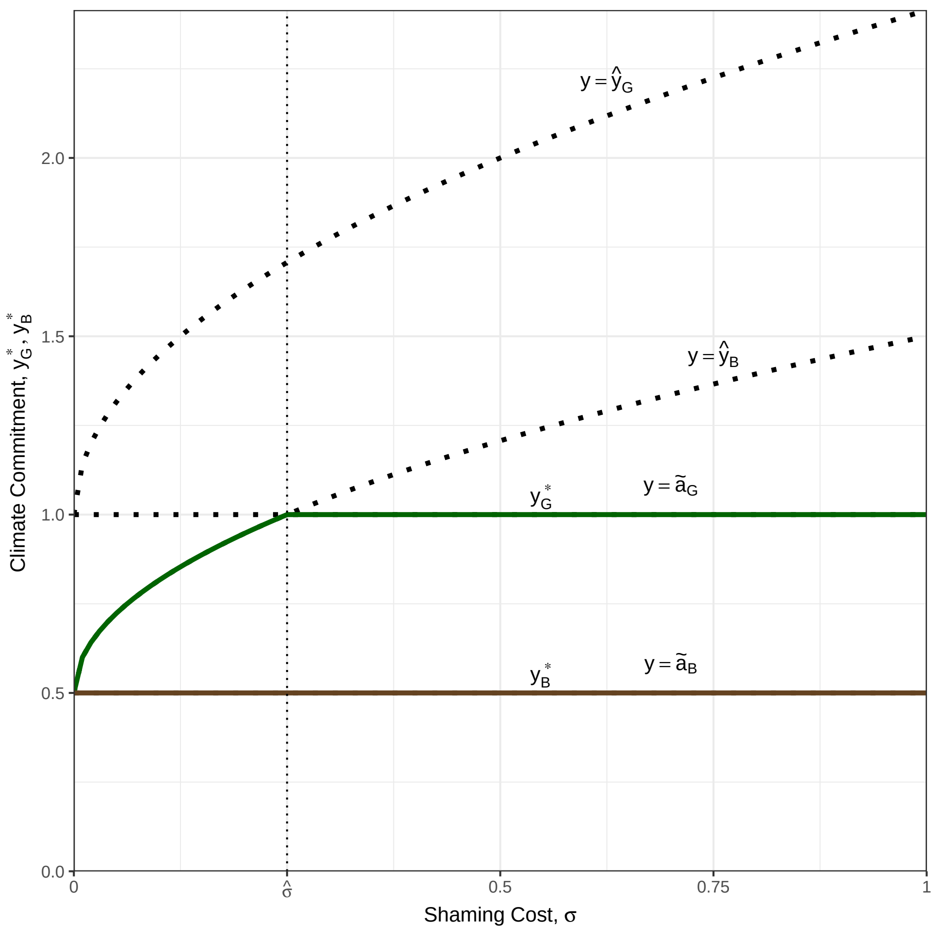
$$y_B^* \leq \tilde{a}_B.$$

Figure 4 illustrates the optimal climate commitments as a function of the shaming cost σ. The green and brown lines plot each party's optimal commitment.

Winning Office

We now examine optimal climate commitments when parties care primarily about winning office, $\Psi \to \infty$. Choosing pledges therefore depends on maximizing the probability of winning the election; G wants to maximize the electoral bias $\Delta(a_G^*, a_B^*; y)$, while B wants to minimize $\Delta(a_G^*, a_B^*; y)$. Figure 5 plots the electoral

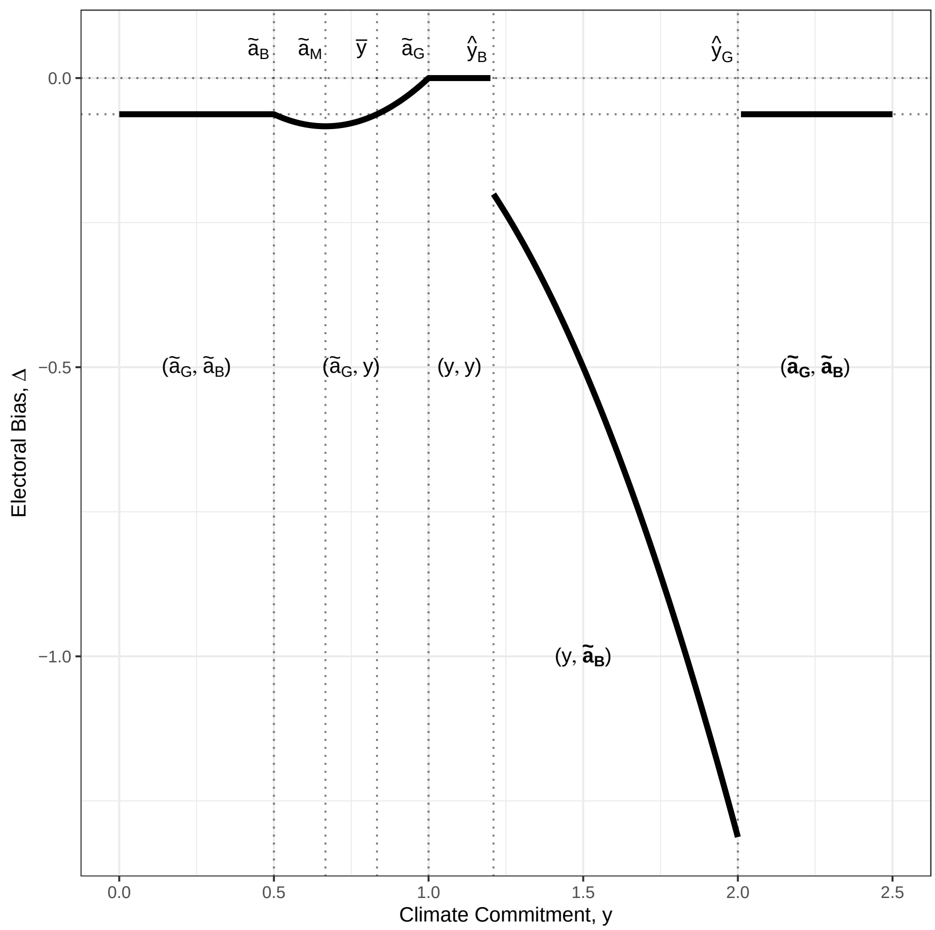
Figure 4: Climate Commitments with Tying Hands Incentives



bias $\Delta(a_G^*, a_B^*; y)$ as a function of the commitment y (for relatively large σ). 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 5. Likewise, if the commitment were very large, $y > \hat{y}_G$, then neither party would attempt to fulfill the commitment, both would implement their ideal points and both would be shamed. Such an overly ambitious target results in the same baseline electoral bias ($\Delta(\tilde{a}_G, \tilde{a}_B; y) < 0$) in B's favor and is shown by the the flat line on the right of Figure 5.

Figure 5: Climate Commitments and Electoral Bias



Next consider the range of commitments between B and G's ideal points: $y \in (\tilde{a}_B, \tilde{a}_G)$. In this region the pledge binds B's policy choice to $a_B^* = y$, while G would implement its ideal point $a_G^* = \tilde{a}_G$. As the pledge initially increases above $\tilde{a}_B$ 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 $a_G^* = \tilde{a}_G$). The electoral bias reaches a local minimum at $y = \tilde{a}_M$ when B implements the median voter's ideal point.

As y 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. We define $\bar{y} = \frac{2\lambda_B - \lambda_M}{\lambda_B \lambda_M}$ as the policy commitment such that, if implemented, the median voter would be indifferent between $\bar{y}$ and B's ideal point, $\tilde{a}_B$ (i.e. $\bar{y} > \tilde{a}_B$ such that $u_M(\tilde{a}_B) = u_M(\bar{y})$). This commitment can conveniently be thought of using a spatial voting analogy as finding the indifference point to the status quo on the other side of a player's ideal point in agenda setting models. We also define $\bar{\sigma} = \frac{2(\lambda_M - \lambda_B)^2}{\lambda_B \lambda_M^2}$ as the smallest shaming cost such that B would implement $\bar{y}$ if elected (i.e. $u_B(\bar{y}) = u_B(\tilde{a}_B) - \bar{\sigma}$).

As y increases toward $\tilde{a}_G$ the election 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 to implement G's ideal point.

The electoral bias jumps downward (i.e. in B's favor) for commitments above $\hat{y}_B$. 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 Δ, while B wants to minimize the electoral bias. The above analysis of Figure 5 provides a simple characterization of G and B's commitments provided the cost of being shamed is sufficiently large. Green would pick $y_G^* = \tilde{a}_G$ that maximizes the electoral bias, $\Delta(\tilde{a}_G, \tilde{a}_G; y) = 0$; Brown would pick $y_B^* = \hat{y}_G$ that minimizes the electoral bias, $\Delta(\hat{y}_G, \tilde{a}_B; y) < 0$. However, such a analysis is valid only when σ is sufficiently large. When the cost of being shamed is relatively small, such large commitments cannot be credibly implemented.

The proposition below specifies the optimal commitments for office-seeking parties for all possible shaming costs. Define $\hat{\hat{\sigma}}$ 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 smallest shame cost such that the largest commitment that G can credibly implement that produces the same electoral bias as B committing to the median voter's ideal effort ($y = \tilde{a}_M$).

Proposition 3 *Let $\beta \to \infty$, $\Psi \to \infty$, and $F' > 0$. G's optimal climate commitment is*

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

B's optimal commitment is

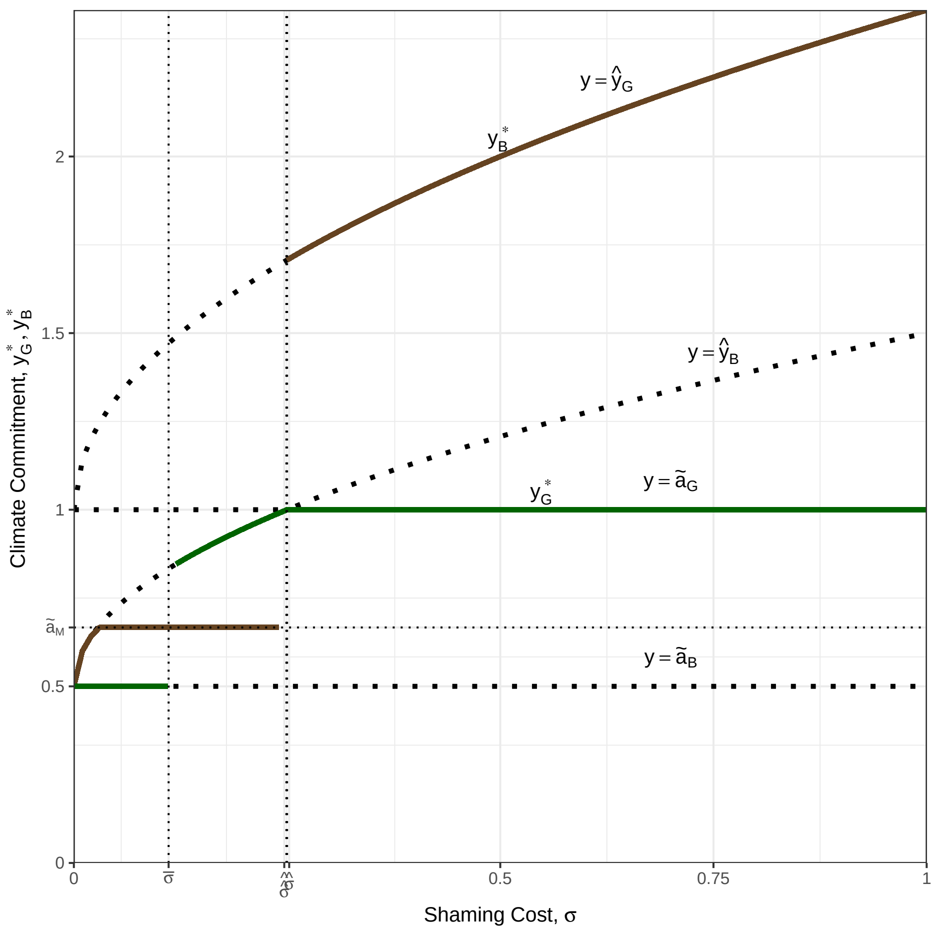
$$y_B^* = \begin{cases} \min\{\tilde{a}_M, \quad \hat{y}_B = \frac{1+\sqrt{2\sigma\lambda_B}}{\lambda_B}\} & \text{if } \sigma < \bar{\sigma} \\ \tilde{a}_M & \text{if } \bar{\sigma} \leq \sigma \leq \hat{\hat{\sigma}} \\ \hat{y}_G = \frac{1+\sqrt{2\sigma\lambda_G}}{\lambda_G} & \text{if } \sigma > \hat{\hat{\sigma}}.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

Figure 6 plots the optimal climate commitments characterized in Proposition 3. A simple discussion of incentives provides proof for the proposition. Starting with the Green party's policy commitment, if the shame cost is low, $\sigma < \bar{\sigma}$, then G can only tie B's hands to implement small efforts over B's ideal point. However, such small increases in the policy that B will implement makes B more electorally attractive to the median voter and so G does not want to pick any policy $y \in (\tilde{a}_B, \bar{y})$ since that reduces the electoral bias. If the shame cost if larger than $\bar{\sigma}$, then G can induce B to implement a policy larger than $\bar{y}$ which moves B's post-election policy further away from the median voter's ideal point (in utility terms) than having B implement its ideal point; this improves G's electability. If the shame cost is sufficiently high, $\sigma \geq \hat{\sigma}$, then the commitment $y_G^* = \tilde{a}_G$ ensures that, post-election, both parties will implement G's ideal point which both removes any electoral bias in favor of B and results in G's most preferred policy whoever is elected.

Turning to Brown's optimal climate commitment when office-holding is the dominant consideration, if the shame cost is small ($\sigma < \bar{\sigma}$), then B wants to make the largest commitment it can adhere to ($\hat{y}_B$) up to the median voter's ideal point. Picking the median voter's ideal point, or the closest point to it that B can commit to, maximizes B's electoral chances.

As the cost of shaming rises further then B has two means of gaining an electoral advantage. First it can continue to commit itself to the median voter's ideal point and have G implement G's ideal point. This approach results in an electoral bias of $\Delta(\tilde{a}_G, \tilde{a}_M; y = \tilde{a}_M)$. Second, B can set a more ambitious climate target, one that it would not adhere to, but one that G will implement. Setting this target as high as is credibly possible, which means $\hat{y}_G$, B induces G to implement policies that are unattractive to

Figure 6: Climate Commitments and Shaming Costs for Office Seeking Parties



the median voter, although B can no longer commit to implementing the median voter's ideal point. This approach induces an electoral bias of $\Delta(\hat{y}_G, \tilde{a}_B; y = \hat{y}_G)$. When $\sigma \geq \hat{\hat{\sigma}}$, B can induce G to implement a sufficiently high level of climate mitigation such that the latter approach induces a smaller electoral bias ($\Delta(\hat{y}_G, \tilde{a}_B; y = \hat{y}_G) \leq \Delta(\tilde{a}_G, \tilde{a}_M; y = \tilde{a}_M)$) and so is preferred by B.

Discussion

We provide a model in which domestic political incentives inform 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 ($\bar{\sigma} < \sigma < \hat{\hat{\sigma}}$) 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*). 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 (cf. Hollyer and Rosendorff 2011). 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

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* .

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

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.".

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.

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.

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.

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. 2023. "Public Opinion Foundations of the Clean Energy Transition.".

Green, Edward J. and Robert H. Porter. 1984. "Noncooperative Collusion under Imperfect Price Information." *Econometrica* 52(1):87–100.

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.

Hollyer, James R. and B. Peter Rosendorff. 2011. "Why Do Authoritarian Regimes Sign the Convention Against Torture? Signaling, Domestic Politics and Non-Compliance." *Quarterly Journal of Political Science* 6(3–4):275–327.

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.

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

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.

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

Melnick, Justin. 2024. "International Climate Change Mitigation, Information Dissemination, and Domestic Politics.".

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

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

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

Nordhaus, William. 2015. "Climate Clubs: Overcoming Free-Riding in International Climate Policy." *American Economic Review* 105(4):1339–1370.

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.

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.

Stokes, Leah C. 2016. "Electoral Backlash against Climate Policy: A Natural Experiment on Retrospective Voting and Local Resistance to Public Policy." *American Journal of Political Science* 60(4):958–974.

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.

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.".

UNDP. 2021. Nationally Determined Contributions (NDC) Global Outlook Report 2021: The State of Climate Ambition. Technical report United Nations.

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

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 $\sigma\sqrt{\beta}\phi\left(\sqrt{\beta}(y - a_g)\right)$ term to the FOC.

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

and the second-order condition (SOC)

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

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 2 *If $\beta < \frac{\sqrt{2e\pi}\lambda_g}{\sigma}$ (or equivalently $\sigma < \frac{\sqrt{2e\pi}\lambda_g}{\beta}$), 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 $a_g^* \geq \lambda_g$.*

Proof of Lemma 2: The second order condition is $soc = -\lambda_g + \sigma\beta\sqrt{\beta}(y - a_g)\phi(\sqrt{\beta}(y - a_g))$, which has a maximum of $\frac{\beta\sigma}{\sqrt{2e\pi}} - \lambda_g$ at $y - a_g = \frac{1}{\sqrt{\beta}}$. Hence if $\beta < \frac{\sqrt{2e\pi}\lambda_g}{\sigma}$ then soc is always negative and the government's optimization is globally concave and foc is decreasing in a_g. At $a_g = \frac{1}{\lambda}$, $foc \geq 0$ 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 $\frac{1+\sqrt{2\lambda_g\sigma}}{\lambda_g}$), 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 3 *1. If $y \leq \frac{1}{\lambda_g}$, then $a_g^* \geq \lambda_g$.*

2. If $y \geq \frac{1+\sqrt{2\lambda_g \sigma}}{\lambda_g}$, 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 $y < \frac{\sqrt{2}\sqrt{\lambda_g}\sqrt{\sigma}+1}{\lambda_g}$; and $a_g^* \to \frac{1}{\lambda_g}$ if $y > \frac{\sqrt{2}\sqrt{\lambda_g}\sqrt{\sigma}+1}{\lambda_g}$.

Proof of Lemma 3: For part 1, if $y \leq \frac{1}{\lambda_g}$, then for $a_g < 1/\lambda_g$ the government payoff are 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 $\frac{1}{2\lambda_g} - \sigma$. 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 $y \geq \frac{1+\sqrt{2\lambda_g\sigma}}{\lambda_g}$. 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, $\lambda_g a_g^* - 1 = \sigma\sqrt{\beta}\phi(\sqrt{\beta}(y - a_g^*))$. 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 $-\lambda_g + \frac{\beta+\beta\lambda_g^2 y^2 - 2\beta\lambda_g y}{4\lambda_g}$. 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 $u_g(a_g) = A - \frac{\lambda_g a_g^2}{2} - \sigma$ 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 $y = \frac{\sqrt{2}\sqrt{\lambda_g}\sqrt{\sigma}+1}{\lambda_g}$ follows directly from equating these payoffs. ■