Pledge-and-Review and Domestic Political Survival

Justin Melnick*

September 15, 2025

Abstract

International organizations increasingly adopt flexible pledge-and-review arrangements based on voluntary commitments and transparency mechanisms with the expectation that information dissemination enhances global public goods provision. However, this paper demonstrates that pledge-and-review mechanisms can have unintended consequences for compliance with international goals. Using a formal model in which leaders simultaneously engage with an international organization and are motivated by domestic electoral incentives, the paper demonstrates that information dissemination does not beget more ambitious commitments. Instead, leaders anticipate how publicized information will affect voter beliefs and, in equilibrium, systematically obfuscate at the pledge stage. As a result, leaders' expected effort toward global public goods provision is no greater in a world with pledge-and-review compared to a world without it. This study highlights a tradeoff between international cooperation and domestic political survival, offering new insights into institutional design and the effectiveness of pledge-and-review frameworks.

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

International organizations (IOs) are often described as agents of global cooperation that can increase public goods provision by providing information and reducing uncertainty (Keohane 1984; Milner 1997; Abbott and Snidal 1998; Fang 2008). In this tradition, IOs are considered neutral arbiters whose role in disseminating credible information facilitates compliance with international commitments (Mitchell 1998). Over time, these organizations have developed increasingly flexible institutional arrangements to serve these purposes, focusing on voluntary national commitments to global public goods and international transparency mechanisms like pledge-and-review frameworks (Harstad 2023*a*;*b*). Herein, political leaders determine the extent to which they wish to contribute to global cooperative efforts, and institutions aggregate and disseminate information about each state's willingness to cooperate. Does information revelation within these frameworks beget compliance with international cooperative goals?

Previous scholarship about the role that IOs play in information dissemination underappreciates the strategic incentives of national leaders within pledge-and-review frameworks, who navigate a dual audience: the international community and their domestic electorate (Mitchell 1998; Haftendorn, Keohane and Wallender 1999; Dai 2002). The same transparency that IOs are designed to promote, by compelling or enabling nations to reveal their plans for compliance with international goals (e.g., Harrison and Lagunoff 2017; Slechten 2020; McAllister and Schnakenberg 2022, in the context of international climate agreements), may impose political costs on leaders at home. In this paper, I show that when international organizations generate publicly available information about leaders' plans to provide global public goods, they create informational spillovers that reshape domestic political incentives to truthfully report and comply with ambitious cooperative objectives. Specifically, leaders anticipate that the information released by IOs may reveal domestically salient details about their preference congruence with the domestic electorate, or whether a leader is beholden to special interests that inhibit the provision of global public goods, which could harm a

leader's political survival.

This paper develops a formal model of how leaders respond to the informational functions of IOs when they interact with leaders' domestic political goals. Inspired by pledge-and-review institutions, I model a setting in which leaders submit public plans detailing their intended contributions to a global public good to an IO, which the IO compiles into a report. For leaders, these disclosures are costless to produce but may be informative about the leader's underlying preferences. Keeping with its traditional role of disseminating information, the IO publicizes the leader's intended contributions. After disclosure, the leader must choose how much costly effort to exert toward the international objective. Crucially, domestic voters observe both the IO's report and the downstream policy effort, and update their beliefs about the leader's willingness to act in the public interest accordingly.

In the model, some "honest" leaders have preferences over public goods provision that align closely with the public interest and wish to reveal this alignment; others are "captured" by special interests and prefer to conceal it. Neither the IO nor the public can distinguish between these types *ex ante*, but the intent of the IO is to "screen" for leaders making ambitious commitments that conform to public demand for global public goods. Domestic voters use the information revealed to update their beliefs about the leader's willingness to enact policies consistent with the public's interest (Fearon 1999; Canes-Wrone, Herron and Shotts 2001; Maskin and Tirole 2004; Acemoglu, Egorov and Sonin 2013). Paradoxically, while IOs have the potential to motivate greater global efforts toward international goals by disseminating information about others' willingness to cooperate (Keohane 1984; Mitchell 2006; Awad and Riquelme 2024), they may also inadvertently deter such efforts when leaders fear that this information will be used by domestic audiences to evaluate their ideological congruence. The downstream domestic politics of signaling thus complicate international efforts to foster cooperation.

The central finding is counterintuitive: leaders obfuscate in their disclosure of intended

contributions precisely because the IO publicizes this information. Rather than demonstrate openness to international cooperative goals through an ambitious pledge *ex ante*, leaders prefer to shift the burden of signaling to downstream policymaking where costly effort, though more expensive, offers a politically safer signal to domestic audiences. In equilibrium, the presence of the IO does not inspire any more effort toward global public goods provision compared to a world without it. Interestingly, the problem does not depend on institutional strength but lies within institutional structure: IOs designed explicitly to disseminate information, and without proper enforcement power, can be strategically subverted by leaders seeking to maintain their domestic political office.

The mechanism driving the result is that leaders make strategic pledges based on how voters might interpret them. The international community prefers ambitious pledges because they could lead to stronger efforts toward global public goods provision downstream. But since ambitious pledges might be informative that the leader shares the public's preferences, leaders who are captured by special interests have incentives to overcommit to ambitious public goods provision but underdeliver in their ultimate efforts (similar to the finding in the international negotiations literature that (generally hawkish) incumbents have incentives to negotiate "weak treaties," see Buisseret and Bernhardt 2018; Battaglini and Harstad 2020; Melnick and Smith 2023). Honest leaders, whose preferences align with the public and who value global public goods more than captured leaders, want to separate themselves from captured leaders, both to implement better policy and to win reelection. However, in equilibrium, a chase-and-evade logic emerges (cf. Roberson 2006): captured and honest leaders make ambitious pledges with the same probability, obscuring the voter's inferences. As a result, pledges are uninformative of the leader's preferences, and they no longer motivate greater effort.

This paper's insight sits at the intersection of two large literatures in international relations: the value of flexibility in institutional design and the role that IOs play as information

disseminators to improve cooperation. Oftentimes international agreements are critiqued for their rigidity if their enforcement regimes are poorly suited to accommodate diverse domestic circumstances (Rosendorff and Milner 2001; Rosendorff 2005; Johns 2014). In response, many contemporary institutions, like the Paris Agreement in the international climate arena, employ pledge-and-review structures. These agreements are celebrated because they are maximally flexible (Linos and Pegram 2016) albeit quite shallow (Downs, Rocke and Barsoom 1998; Gilligan 2004; Edry 2020). Typically, they do not have formal enforcement mechanisms beyond the dissemination of information about intended contributions to public goods.

Proponents of pledge-and-review institutions argue that flexibility enhances buy-in without sacrificing membership (Falkner 2016; Keohane and Oppenheimer 2016). The "pledge" recenters domestic politics in the development of international cooperative goals by providing leaders with total control over their commitments to public goods provision. The international institution then publicizes pledges and facilitates "review" via information sharing, but not does not have explicit power to compel nations to meet their pledges. Taken together, these features constitute a "bottom-up" agreement, compared to an arrangement where an IO imposes "top-down" targets that are legally enforced, which should entice a larger number of states to participate and generate greater aggregate welfare gains (Harstad 2023*b*). However, more circumspect observers of pledge-and-review question the institution's ability to alter cooperative depth (Bang, Hovi and Skodvin 2016). In this vein, I argue that the flexibility of commitments interacts with a leader's domestic political circumstances in a counterproductive way: leaders are incentivized to obfuscate in their pledgemaking, sacrificing the informativeness of international organizations if they anticipate that such information dissemination could be electorally harmful. It is precisely the recentering of domestic politics that defangs the power of pledges, as in equilibrium the institution does not induce costly behavioral change.

This paper is the first to propose a theoretical analysis of pledge-and-review-style frameworks that integrates goals of international compliance with domestic political incentives. The formal literature on such institutions is scant. Harstad (2023*a*;*b*) advances a series of papers to study optimal commitments but does not consider the role that domestic politics plays in leaders' incentives to forge commitments. A related paper is Melnick and Smith (2025), which studies the domestic consequences of the Paris Agreement on electoral competition, but does not consider the institution's role in disseminating information. I address this gap by explicitly considering how leaders' incentives to cooperate interact with institutional features of information dissemination in front of a domestic audience that may find such information electorally relevant. Since proponents of these agreements argue that the recentering of domestic politics is crucial to their success (Falkner 2016), we need a model in which leaders make voluntary pledges, these pledges are publicized, and the publicization of these pledges interacts with the domestic political environment. As we shall see, once this is taken into account, the primary benefit of pledge-and-review is nullified as leaders' optimal pledges induce effort that is identical to a world without the agreement.

More broadly, I highlight a novel effect of domestic politics within the literature on information and cooperation. Many of these theories treat domestic politics as an enforcement mechanism to enhance accountability: IOs provide information to voters about whether a leader has defected from cooperative agreements, and voters can punish defectors (Mansfield, Milner and Rosendorff 2002; Dai 2005; Fang 2008; Chaudoin 2014). This is done without modeling the leader's strategic disclosure of information to be disseminated by IOs. Here, by contrast, I model both an opportunity for leaders to document their willingness to comply with international goals as well as the stage in which they exert effort toward those objectives, and domestic accountability affects both stages. Domestic politics are not merely a constraint on compliance but a central driver of what an IO knows, based on the reports it receives from leaders, and can disseminate to the public (e.g., Bayer and Crippa 2025). It

is precisely the informational effect that the agreement has on domestic political dynamics that stifles leaders' intentions of committing to bold pledges in the first place.

This argument also builds on recent research examining the domestic political effects of international benchmarking and transparency. Institutions like the World Bank, OECD, and United Nations routinely issue performance reports designed to motivate better governance by exposing underperformance or highlighting best practices (Kelley and Simmons 2015; 2019). One may expect that greater information provision will lead to better governance because information should strengthen the accountability channel between leaders and their publics. On the other hand, the publicization of information by IOs may have unintended consequences by enticing leaders to strategically substitute effort away from other priorities toward meeting the published goal (Bisbee et al. 2019). This paper highlights how the strategic disclosure of information to IOs, and its subsequent publicization to domestic audiences, may mediate the relationship between information dissemination and improved governance outcomes. Relatedly, studies on transparency demonstrate that information disclosed by leaders, often filtered through IOs, affects voter beliefs and subsequent political support for the incumbent (Hollyer, Rosendorff and Vreeland 2015; 2018; 2019*a*;*b*). Defined as the precision with which a citizenry knows economic fundamentals, transparency can breed mass instability in autocratic regimes but stabilizes democratic regimes. My argument is that, within institutions like pledge-and-review frameworks, leaders may be able to alter the effects of information dissemination in order to preserve their domestic survival.

This model complements but departs from existing work on how leaders and nations shape the functioning of international institutions to serve their needs. Canonical scholarship considers how powerful states design institutions with the intent of entrenching their interests within a world order (Gruber 2000; Stone 2011; Lipscy 2017). Similarly, this paper studies how domestic interests shape leaders' engagement with international institutions and shows how international goals may be subverted to accommodate political survival concerns.

Melnick and Smith (2025) argue that leaders may set optimal climate commitments to vie against domestic political competitors, only setting ambitious targets if it helps leaders win elections. Fang (2008) shows that when leaders choose whether to consult an IO, the act of consultation itself serves as a signal of their preferences. Here, however, leaders are already members of the IO and must decide how to respond to informational requests. The signal is not whether they engage but what they reveal and how they follow through. This distinction shifts the analytic focus from the decision to comply to the downstream consequences of information dissemination.

Empirically, the model helps to make sense of the inconclusiveness of studies examining the effects of nationally determined contributions on efforts to provide public goods. In the case of climate commitments under the Paris Agreement, Tørstad and Wiborg (2024) find that more ambiguous pledges are less ambitious. Ambiguity is consistent with the equilibrium outcome of the model, in which the IO, and subsequently publics, are unable to ascertain leaders' preferences for public goods provision from their commitments. In the climate space, this often manifests as the omission of technical details that leave commitments open to interpretation (Rogelj et al. 2017; Pauw et al. 2018; Rowan 2019), which makes it difficult to assess whether commitments are credible (Nemet et al. 2017; Elliott, Bernstein and Hoffmann 2022; Victor, Lumkowsky and Dannenberg 2022; Dannenberg et al. 2023), thereby generating mixed empirical findings.

Empirical Referent: Global Climate Change Governance

In what follows, it may be useful to consider the contemporary structure of international climate change governance as an empirical referent, and the mechanism this governance network relies on to motivate mitigation efforts. While I do not seek to model the international politics of climate change *per se*, this empirical grounding is useful to think about

the institutional foundations that undergird international information dissemination within pledge-and-review systems, like those that govern international climate policymaking.

The contemporary international climate change "regime complex" comprises a patchwork of institutions that are loosely coupled but share the goal of combating climate change by encouraging mitigation and adaptation investments (Keohane and Victor 2011). Two primary organs within this vast network are the Intergovernmental Panel on Climate Change (IPCC) and the Paris Agreement. The IPCC, which organizes the most comprehensive, global assessment of scientific evidence on climate change (De Pryck and Hulme 2022; Hughes 2024; Crippa and Bayer 2025), supports the work of the United Nations Framework Convention on Climate Change (UNFCCC), which led to the negotiation of the Paris Agreement in 2015. The IPCC provides the scientific foundation for the Paris Agreement by assessing climate change and its impacts. The Paris Agreement, in turn, builds upon the IPCC's scientific findings and evaluations to provide a framework that charts the political goals for global climate action.

These two institutions together serve an important role in aggregating and disseminating information about the science surrounding the impacts of climate change and the political response to it. Notably, they are both *advisory bodies* (Fang and Stone 2012) that benefit from substantial input and buy-in from national governments (Bayer and Crippa 2025): the IPCC's primary task is to assess and summarize existing knowledge rather than conduct its own research, providing summaries for policymakers that communicate key findings via public statements. Given public knowledge and national willingness, the Paris Agreement requires that political leaders submit nationally determined contributions (NDCs), plans that document a nation's expected but voluntary emissions reductions policies, which are then publicized.[1] NDCs are intended to capture nations' willingness to execute on international

[1]Institutions with similar structures in other issue areas include the OECD's Development Assistance Committee Peer Reviews, the UN Convention Against Corruption's Implementation Review Mechanism, the World Trade Organization's Trade Policy Review Mechanism, the Universal Periodic Review of the

climate goals as a function of domestic flexibility and feasibility (Keohane and Oppenheimer 2016; Falkner 2016).

On the back end, the Paris Agreement has no formal means of inducing greater mitigation commitments from leaders. It is legally endowed with its review process to complement pledging, but strict enforcement is not codified. Policymakers anticipate informal reputational sanctions, or "naming and shaming," to be levied either by the international community or by domestic publics if leaders are found to transgress their commitments (Hafner-Burton 2008; Tingley and Tomz 2022; Casler, Clark and Zucker 2023; Melnick and Smith 2025).

Paris's formal review process is multifaceted. The agreement's "global stocktake" provides additional information about nations' performance in meeting their NDCs *ex post*, as described in Article 7.14 (UNFCCC 2015). Articles 13, 14, and 15, develop an "enhanced transparency framework" for information dissemination that motivates compliance with goals to limit global warming. This framework is expected to help track progress toward countries' NDCs, as parties are required to submit biennial transparency reports every two years, which then provide a foundation for future global stocktakes. Additionally, the streamlining of the transparency process intends to narrow the gap between developed and developing nations that previously existed under erstwhile climate governance institutions: the enhanced transparency framework requires all countries to submit their progress reports on the same timeline for a common evaluation process (Konrad, van Deursen and Gupta 2022).

The role of information is therefore central in the climate regime complex, but its implementation is in nascent stage. The body recently received its first transparency reports in December 2024. While little is known about the downstream effects of its implementation,

Human Rights Council, and the International Labor Organization's Complaint Procedure (Raiser, Çalı and Flachsland 2022).

scholars hope that the transparency gains that exist on paper translate into practice. Kamil and Karlsson-Vinkhuyzen (2023) argue through case studies of five Global South countries that policymakers are aware of the implications of transparency and the dissemination of mitigation-related information on expected country performance toward their climate goals. Raiser et al. (2020) write that while the Paris Agreement's transparency system is well-designed structurally, real-world effectiveness is expected to be mixed: transparency has improved reporting visibility but has not reliably translated into stronger mitigation or more ambitious NDCs. Other scholars question the value of enhanced reporting more directly, as these institutions may constrain countries' capacities to autonomously form their NDCs (Konrad, van Deursen and Gupta 2022; van Deursen and Gupta 2025).

This institutional setup poses an interesting political problem for national leaders, who must determine whether to truthfully disclose their willingness and ability to tackle climate change and how the publicization of such commitments shapes domestic politics. The model is thus useful as a fairly general framework in guiding what we might expect in terms of how transparency mechanisms interact with leaders' incentives to make ambitious commitments. As discussed above, future information revelation is paramount to the functioning of pledge-and-review frameworks like the Paris Agreement. The initial NDCs that leaders pledge form the basis of any inputs to compliance assessments. As we shall see, the equilibrium effects of this anticipated information dissemination validate scholars questioning the practical efficacy of transparency frameworks on NDC ambition.

Motivated by this empirical referent, the theory presents an abstract representation of an IO which captures the essential features information aggregation and dissemination within a model of domestic policymaking.

Model

The model embeds domestic and international policymaking in which $i = 1, \ldots, n$ countries interact over two periods. I contrast two cases: in the first, nations act independently in allocating costly effort toward providing global public goods (e.g., environmental remediation), and in the second there is an international organization (IO) that disseminates information about nations' intended policy choices. Throughout I focus on domestic decisionmaking within a representative nation.

Domestic Politics

Within the representative nation there is a leader (she) and a representative, or median, voter (he).[2] Leaders determine how much effort to exert toward global public goods provision in each period, $a_{i,t} \geq 0$. In between periods, there is an election in which the voter determines whether to retain or replace the leader.

Leaders, as members of the international community, value global efforts as well as remaining in domestic office. Effort is globally beneficial, but also locally costly to implement. I assume that leaders vary in the extent to which they are influenced by special interests, which affects their willingness to exert effort into global cooperation. I parameterize this willingness through leaders' marginal costs of effort. "Honest" leaders have preferences congruent with the median voter and produce effort at marginal cost $\lambda_M > 0$; conversely, "captured" leaders also act on behalf of anti-reform special interest groups and thus have greater marginal costs of effort production, $\lambda_M + \lambda_S$. We can think of this as fossil fuel or other anti-environmental industries seeking to steer leaders away from environmental regulations, raising their marginal costs of policy reform. While not explicitly modeled, greater

[2]The language of voters and elections is used to ease exposition, but the model need not be scoped to democratic countries. In nondemocratic states, this player may be an elite or other individual whose political support is pivotal for the leader's survival in office (Bueno de Mesquita et al. 2003).

marginal costs for captured leaders align with the opportunity cost of foregone campaign contributions from interest groups (Grossman and Helpman 1994; Acemoglu, Egorov and Sonin 2013; Kennard 2020). Whether a leader is honest or captured is their private information; the voter believes that the incumbent leader is honest with probability $\gamma \in (0,1)$. Let $\theta \in \{0,1\}$ be an indicator of the leader's type, where $\theta = 1$ denotes an honest leader and $\theta = 0$ denotes a captured leader.

Policies that finance global public goods impose costs on domestic publics. It is often voters who have to pay for costly climate reforms (e.g., Bechtel and Scheve 2013; Ansolabehere and Konisky 2014; Stokes 2016; Gazmararian and Tingley 2023; Voeten 2025), where the median voter's willingness to pay is summarized by his marginal cost λ_M. The voter, who has preferences over global efforts, has a per-period payoff written as (suppressing dependence on i and t; as we shall see, most of the analysis is focused on the first period)

$$u_M(a, A; \lambda_M) = A - \frac{\lambda_M}{2} a^2,$$

where $A = \sum_i a_i$ is the per-period global effort exerted toward international cooperation. The first term of this utility function captures the fact that everyone is better off when leaders across the globe exert more effort toward providing global public goods. The second term captures the local costs: for example, voters may be shouldered with costly climate reforms in order to abate emissions within nation i. It is clear that the median voter's ideal level of effort toward global public goods provision is $\frac{1}{\lambda_M}$: a smaller λ_M means that the median voter in nation i prefers greater effort toward public goods provision.

The incumbent leader's utility involves a policy payoff, reflecting either the median voter's or special interests' preferences over global public goods provision, and a benefit from holding office.[3] Without loss of generality, leaders obtain utility normalized to zero when not in office.

[3]This utility function, standard in the literature (e.g., Harstad 2012; Melnick and Smith 2025), keeps analysis tractable across countries and allows for the isolation of the domestic political mechanism.

Their per-period utility in office can be expressed as

$$v_L(a;\theta) = \beta\Big(A - \theta\frac{\lambda_M}{2}a^2 - (1-\theta)\frac{\lambda_M+\lambda_S}{2}a^2\Big) + \Psi,$$

where $\beta > 0$ scales the leader's weight between voter welfare and office-holding and $\Psi > 0$ is the benefit of office-holding. The utility function demonstrates that the honest leader's policy preferences are aligned with those of the median voter, while the captured leader's policy preferences are biased by the expected costs of effort imposed on special interest groups. These differential marginal costs imply differences in ideal points:[4] the captured leader has ideal effort $\frac{1}{\lambda_S+\lambda_M} < \frac{1}{\lambda_M}$, meaning the special interest's position—and by extension that of the captured leader—is unpopular with at least a majority of voters.

The leader's effort choice is imperfectly observed by the voter, who observes the noisy signal $x = a + \varepsilon$, where $\varepsilon \sim N(0, \frac{1}{\zeta})$ and $\zeta < \frac{\lambda_M\sqrt{2\pi e}}{\Psi}$. We can think of ζ as domestic "transparency" or "bureaucratic quality" (Slough 2024), the precision with which the voter perceives the mapping from the leader's effort to policy outputs. Uncertainty in the mapping from effort to policy outcomes may also represent leaders' inability to forecast how policy may affect a pivotal voter (Gazmararian and Tingley 2023). The policy outcome is informative about the types of policies that leaders will implement in the future (second period), as higher values of the signal are more likely to reflect more ambitious effort levels. The voter casts a ballot either for the incumbent leader or a potential challenger based on the realization of x. If the incumbent loses the election, the challenger assumes office and is honest with probability γ.

[4]This parameterization is isomorphic to leaders and voters have quadratic policy preferences over effort where the honest leader and the voters have utility $-(a - \frac{1}{\lambda_M})^2$ and the captured leader has utility $-(a - \frac{1}{\lambda_M+\lambda_S})^2$.

International Politics

In an augmented version of the game, leaders engage through an international organization (IO) prior to the exertion of effort. Leaders first send a costless message $m \in \{0, 1\}$ to the IO, which represents a plan about expected downstream effort by reporting a leader's willingness to provide global public goods. In this sense, a message of leader type corresponds to the extent to which leaders would be willing to take on costs of cooperation given their relationship with the public. Without loss of generality, a message $m = 1$ can be interpreted as the leader submitting an optimistic plan about global public goods provision to the IO, consistent with the preferences of the median voter. By contrast, a message $m = 0$ implies that a leader is willing to exert less effort as such a leader is beholden to special interests.

The message m is reported to the IO who then produces a report. In pledge-and-review frameworks, we can think about this as the IO aggregating and disseminating information about leaders' nationally determined contributions[5] (Falkner 2016; Keohane and Oppenheimer 2016), which may facilitate a global stocktake or other form of review process (Melnick and Smith 2025). As a core function of IOs is information dissemination, IOs publicize leaders' willingness to cooperate. The report takes the form $s = m + \nu$ where $\nu \sim N(0, \frac{1}{\tau})$. The parameter τ is thus a measure of institutional strength: larger τ means that the IO's report of the leader's type is more precise, and therefore the report s is more informative. The report s is observed by voters as an additional piece of information about the leader's type. The IO is not a strategic player, but this formalization represents the idea that it is a neutral arbiter of information intending to maximize effort (although see Fang and Stone 2012).

The timing of the game is as follows (where stages 1 and 2 below are only played in the

[5]Substantively, national determined contributions can be a broad pledge about a leader's willingness to tackle climate change—spanning mitigation commitments, adaptation, technological investments, reporting requirements, or the creation of data inventories. This flexibility allows countries to pledge more than emissions reductions; the model presupposes that downstream effort provides a global public good, so pledges should be thought of primarily as relating to countries' mitigation commitments to facilitate interpretation.

model with international politics).

1. The incumbent leader sends message $m_1 \in \{0, 1\}$ to the IO.
2. The IO disseminates report $s_1 = m_1 + \nu_1$ to the public.
3. The incumbent leader exerts effort $a_1 \in \mathbb{R}_+$.
4. The voter observes the signal $x_1 = a_1 + \varepsilon_1$ and decides whether to retain or replace the incumbent leader.
5. The winner of the election sends message $m_2 \in \{0, 1\}$ and exerts effort $a_2 \in \mathbb{R}_+$.
6. Payoffs realized. Game ends.

The solution concept is weak Perfect Bayesian equilibrium. A strategy for the leader constitutes a choice of (message m and) effort a given type θ (and the IO's report s). A strategy for the voter maps the signal x (and IO report s) into a reelection decision. Voter belief's about the leader's type are formed via Bayes's rule.

Comments on the Model

Before turning to analysis, there are a few model assumptions that require further comment.

As described above, information sharing is an integral part of the pledge-and-review process. The IO in this model is instantiating the institutional intention to aggregate and disseminate information about intended contributions to global public goods. This function can be interpreted as a means of coordinating effort across countries (Awad and Riquelme 2024) with the objective of providing a common pool of knowledge and expertise off of which nations can build (Fang 2008). Such information sharing can also lead to the "orchestration" of best practices (Hale and Roger 2014; Abbott et al. 2015). This repository of information may also facilitate downstream "naming and shaming" by interested parties who observe the

IO's report and leaders' ultimate efforts to meet their commitments (Milkoreit and Haapala 2017). However, as we shall see, efforts to be coordinated or best practices to be shared, as outputs of the IO, are strategically determined by inputs from the leader who anticipates these reports.

Formally, I operationalize this information sharing process as a report, which is a noisy signal of the leader's message. The message is analogous to the nationally determined contribution, which IOs then publicize. As IOs could aggregate the plans of many nations, the report need not be perfectly precise signals of leaders' messages, captured by the noise term. From a technical perspective, the model is also cleaner by admitting a continuous report s rather than having the voter observe the message m directly because it ensures nondegenerate beliefs about the leader's type on the equilibrium path. This allows for an easier characterization of the equilibrium without appealing to equilibrium refinements.

Beyond international cooperation, leaders also communicate directly with their domestic publics. While IOs may be seen as a more credible means of communicating information (Grigorescu 2003), both forms of communication can be considered cheap talk. In the model, the IO is simply a vehicle for the publicization of the leader's pledge. It is not unreasonable to assume that other cheap talk messages exist, but I leave them unmodeled as the paper does not seek to explain all forms of engagement between leaders and their publics.[6] Within the current framework, we can think about the effects of other messages as incorporated into the prior γ.

[6]It is not obvious why the presence or absence of direct public messaging should affect a leader's optimal commitment within a pledge-and-review system—that is, why a leader's optimal pledging strategy might differ because of any public communication—and thus by not directly modeling it, it is effectively held fixed (Paine and Tyson 2020).

Analysis

I begin by studying the equilibrium of the domestic interaction, and then build to the case where nations announce their intentions for global public goods provision through the IO. Throughout, I first provide statements of formal results and then provide intuition in subsequent discussion. All formal proofs are relegated to the appendix.

Domestic Politics

As a benchmark, I solve the game when leaders act independently and do not submit messages to international organizations. The features of the equilibrium are reminiscent of other classical models of electoral accountability (e.g., Acemoglu, Egorov and Sonin 2013).

Proposition 1. *In the unique Perfect Bayesian equilibrium without the international organization, the first period efforts a_θ^* solve*

$$\underbrace{\beta}_{\text{marginal policy benefit}} + \underbrace{\sqrt{\zeta}\phi(\sqrt{\zeta}(\frac{a_1^* - a_0^*}{2}))\Psi}_{\text{marginal electoral benefit/signaling premium}} = \underbrace{\beta\theta\lambda_M a_1^* + \beta(1-\theta)(\lambda_M + \lambda_S)a_0^*}_{\text{marginal policy costs}}.$$

The voter reelects the incumbent if and only if $x \geq \hat{x}(a^*)$. In the second period, leaders exert effort at their ideal points, $a_\theta^* = \frac{1}{\lambda_M + (1-\theta)\lambda_S}$.

In the second period, the leader does not have to worry about how her choice of effort influences her reelection decision or future policy outcomes. The optimal effort is thus her ideal point, $a_\theta^* = \frac{1}{\lambda_M + (1-\theta)\lambda_S}$. Honest leaders choose the level of effort that is socially optimal within the electorate, exemplified by the median voter's ideal point, while captured leaders are biased against the median voter given their affiliation with special interests. As $\frac{1}{\lambda_M + \lambda_S} < \frac{1}{\lambda_M}$, captured leaders exert less effort into public goods provision than honest leaders because they internalize the preferences of a pro-status quo interest group. It should

be noted that a captured leader therefore does not impose as many costs on the voter because he exerts less effort, but of course this means that he does not provide as many benefits either.

From the voter's perspective, he knows that an honest leader will on average enact policies in the second period that are more in line with their preferences. This selection problem drives the voter's electoral considerations: in equilibrium, the voter's task is to retain honest leaders and replace captured leaders. The voter learns about the leader's type through the realized policy outcome x, which is a noisy signal of the incumbent leader's effort toward global public goods. Let $\mu(x) = P(\theta = 1|x)$ be the voter's posterior belief that the incumbent is honest given the realized policy outcome x and $\hat{a}_\theta$ be the voter's conjecture about the leader's effort (since it is not directly observed). By Bayes's rule,

$$\mu(x) = \frac{\gamma\phi(\sqrt{\zeta}(x - \hat{a}_1))}{\gamma\phi(\sqrt{\zeta}(x - \hat{a}_1)) + (1 - \gamma)\phi(\sqrt{\zeta}(x - \hat{a}_0))}.$$

As the voter eyes the future, he retains the incumbent if and only if his posterior belief about the incumbent's type is greater than the prior probability that a challenger is an honest type, $\mu(x) \geq \gamma$, or if the policy outcome x is greater than some threshold,

$$x \geq \frac{\hat{a}_1 + \hat{a}_0}{2} \equiv \hat{x}(\hat{a}).$$

Given the voter's threshold and conjectures about incumbent effort, the incumbent's reelection probability is equal to

$$\pi(a) = \Phi(\sqrt{\zeta}(a - \hat{x}(\hat{a}))).$$

In the first period, the leader needs to balance her policy preferences, either representing the median voter or special interest groups, with her desire to be reelected. Her maximization

problem is

$$\max_{a} \ \beta\Big(a - \theta\frac{\lambda_M}{2}a^2 - (1-\theta)\frac{\lambda_M+\lambda_S}{2}a^2\Big) + \pi(a)\Psi,$$

such that equilibrium efforts a^*_θ are characterized by:

$$\underbrace{\beta}_{\text{marginal policy benefit}} + \underbrace{\sqrt{\zeta}\phi(\sqrt{\zeta}(\frac{a^*_1 - a^*_0}{2}))\Psi}_{\text{marginal electoral benefit/signaling premium}} = \underbrace{\beta\theta\lambda_M a^*_1 + \beta(1-\theta)(\lambda_M+\lambda_S)a^*_0}_{\text{marginal policy costs}}.$$

Intuitively, since honest leaders have lower marginal costs of effort, and are more aligned with the preferences of the median voter, they exert greater effort than captured leaders in equilibrium, $a^*_1 > a^*_0$. To determine equilibrium effort levels, leaders balance the marginal costs of public goods efforts with two forces. Leaders care about the policy benefits from contributing to the global public good, and leaders exert effort as a means of surviving in office by signaling their type to the voter. Indeed, relative to each leader's ideal point, equilibrium efforts feature a "signaling premium" that goes above and beyond what each leader would prefer on pure policy grounds in order to optimally convince the voter that the leader is honest. As holding office becomes more valuable, the extent of the signaling premium grows.

The signaling premium is essential to the logic of the equilibrium. When the voter observes a higher value of the policy outcome x, he is more likely to believe that the leader is honest, as honest leaders can exert effort more cheaply, which on average translates into better policy outcomes. Better policy outcomes lead to greater odds of reelection. This incentivizes both leaders to commit to greater levels of effort than their ideal points. Unfortunately, such signaling can impose greater costs on voters: even though an honest leader has identical policy preferences as the median voter, she nevertheless overcommits to a more ambitious effort level than the median voter's ideal point, $a^*_1 > \frac{1}{\lambda_M}$, in order to insulate herself from the possibility that the voter perceives her as captured. Similarly, captured leaders

exert effort $a_0^* > \frac{1}{\lambda_M+\lambda_S}$, but need not overburden the voter with the costs of financing public goods; if the special interest's bias is relatively large then the captured leader will choose effort $a_0^* \in (\frac{1}{\lambda_M+\lambda_S}, \frac{1}{\lambda_M})$.

The leader's incentive to signal honesty to the voter implies that, going into the election, the honest leader overburdens the voter with additional costs by exerting greater effort than his ideal point, but ensures better policy outcomes than the captured leader on average in the future. But captured leaders may implement effort that imposes fewer costs on the voter, thereby making them appear electorally attractive. Moreover, relative to the first period, in which leaders "overcommit" to signal honesty, they "underdeliver" in the second period in the sense that policy outcomes are on average lower in the second period relative to the first period.

Straightforward comparative statics follow from the equilibrium. The value of holding office Ψ represents the purpose of the signaling premium, and so increasing the value of reelection implies that leaders will exert greater first period effort. By contrast, increasing marginal costs—which can be construed as souring public opinion on public goods provision for the median voter, increasing λ_M, as well as increased expected compliance costs for special interests λ_S—reduces leaders' incentives to provide public goods. The effect of domestic transparency ζ is non-monotonic because at very high or very low levels of ζ the leader's reelection chances are poorly conditioned on policy outcomes, sending effort to leaders' ideal points; effort is thus greatest when ζ is intermediate.

Three essential yet intuitive results follow from the characterization of the equilibrium and summarize the discussion above.

Corollary 1. *In equilibrium:*

- *honest leaders exert more effort than captured leaders,* $a_1^* > a_0^*$*;*
- *honest leaders exert greater effort than the median voter's ideal point,* $a_1^* > \frac{1}{\lambda_M}$*;*

- *captured leaders exert greater effort than the median voter's ideal point,* $a_0^* > \frac{1}{\lambda_M}$*, if* $\lambda_S < \lambda_M + \sqrt{\frac{\zeta}{2\pi}}\frac{\Psi}{\beta}$.

Corollary 2. *Honest leaders win reelection more than captured leaders,* $\pi(a_1^*) > \pi(a_0^*)$.

Corollary 3. *Leaders' first period equilibrium efforts* a_θ^* *are:*

- *increasing in* Ψ*;*
- *decreasing in* λ_M*;*
- *decreasing in* λ_S*;*
- *inverse-U shaped in* ζ.

This analysis demonstrates how leaders' signaling concerns can incentivize greater effort into global public goods provision.[7] The logic follows from the fact that voters expect honest leaders to deliver better policy outcomes on average than captured leaders, and thus leaders have an incentive to demonstrate their honesty.

In the case of climate change remediation, the equilibrium argument implies that voters are more likely to reward leaders who engage in greater mitigation provision. While empirical support for electoral returns from mitigation is mixed (e.g., Stokes 2016; Bolet, Green and González-Eguino 2024), and some scholarship discusses the possibility of a "green backlash" (Heddesheimer, Hilbig and Voeten 2024; Voeten 2025), the argument is a reasonable approximation for two interrelated reasons. First, public opinion consistently shows publics across the world value mitigation and find tackling climate change to be an important problem (Gazmararian, Mildenberger and Tingley 2025; Dechezleprêtre et al. 2025). Second, despite

[7]The model also provides an explanation for why unilateral investments into public goods may be rational in the face of global collective action problems (see discussions in climate policy Aklin and Mildenberger 2020; Kennard and Schnakenberg 2023). If domestic publics can learn about leader quality from policy implementation, signaling the opportunity to invest in public goods efforts is valuable for leaders beyond free-riding concerns.

the rise of anti-climate populist parties, such political forces have yet to capture an electoral majority; all that is required for the theoretical logic to hold is that voters prefer more effort into public goods than anti-reform special interests. More technically, results are robust to the inclusion of a valence shock which would represent the value of the incumbent leader on all other issues besides investments into global public goods; this weakens the relationship between effort and electoral support but maintains the thrust of the argument (see Proposition A.1 in the appendix).

International Politics

Now consider the augmented model in which leaders submit plans about their intentions to provide global public goods to an international organization. The organization publicizes this information, i.e., to the voter, in the form of a report. In the second period, leaders continue to implement their ideal points, and the message m is not payoff relevant, so I focus on analysis of the first period.

Define $\hat{p}_\theta = P(m = 1|\theta)$ as the voter's conjecture about the probability that leader-type θ sent message $m = 1$ to the IO. The next proposition, similar in form to Proposition 1, incorporates the IO's report into optimal effort and reelection behavior.

Proposition 2. *In the unique Perfect Bayesian equilibrium, the first period efforts* $a_1^*(\hat{p})$ *and* $a_0^*(\hat{p})$ *solve*

$$\underbrace{\beta}_{\textit{marginal policy benefit}} + \underbrace{\sqrt{\zeta}\phi(\sqrt{\zeta}(a_\theta^* - \hat{x}(a^*, \hat{p})))\Psi}_{\textit{signaling premium given IO}} = \underbrace{\beta\theta\lambda_M a_1^* + \beta(1-\theta)(\lambda_M + \lambda_S)a_0^*}_{\textit{marginal policy costs}}.$$

The voter reelects the incumbent if and only if $x \geq \hat{x}(a^*, \hat{p})$. *In the second period, leaders exert effort at their ideal points,* $a_\theta^* = \frac{1}{\lambda_M + (1-\theta)\lambda_S}$.

The proposition clarifies that the IO's report impacts effort through its effect on the

voter's belief that the leader is honest. To see this, denote $\mu(x, s)$ as the voter's posterior belief about the leader's type having observed IO report s and signal x of the leader's effort and $\hat{m}_\theta = \hat{p}_\theta\phi(\sqrt{\tau}(s-1)) + (1-\hat{p}_\theta)\phi(\sqrt{\tau}s)$ as the total probability of s. Given voter conjectures $\hat{a}_\theta$ and $\hat{p}_\theta$, the voter's posterior can be expressed as

$$\mu(x,s) = \frac{\gamma\phi(\sqrt{\zeta}(x-\hat{a}_1))\hat{m}_1}{\gamma\phi(\sqrt{\zeta}(x-\hat{a}_1))\hat{m}_1 + (1-\gamma)\phi(\sqrt{\zeta}(x-\hat{a}_0))\hat{m}_0},$$

such that it is optimal to retain the incumbent leader whenever

$$x \geq \frac{\hat{a}_1 + \hat{a}_0}{2} + \frac{\log(\frac{\hat{m}_0}{\hat{m}_1})}{\zeta(\hat{a}_1 - \hat{a}_0)} \equiv \hat{x}(\hat{a}, \hat{p}).$$

The first term of the cutoff $\hat{x}(\hat{a}, \hat{p})$ is identical to the first model, $\hat{x}(\hat{a})$, but the second term factors in the role of the IO's report into the voter's beliefs about the leader's type. Since the voter now has two pieces of information about the leader, the IO's report and the policy outcome, he assesses outcomes differently. If the voter believes that leaders send $m = 1$ with equal probability, $\hat{p}_1 = \hat{p}_0$, then the message is uninformative and the second term collapses, leaving only the voter's assessment of downstream effort to determine reelection probabilities. In such cases the IO's report is not relevant for the voter's beliefs about the leader's type, and thus not relevant for the leader's effort choice. However, if the voter believes that leader messages may have been informative about type, then the IO's report sets the leniency of the voter's reelection cutoff.

Corollary 4. *If $\hat{p}_1 \neq \hat{p}_0$, the voter's threshold $\hat{x}(\hat{a}, \hat{p})$:*

- *increases in s if $\hat{p}_0 > \hat{p}_1$;*
- *decreases in s if $\hat{p}_1 > \hat{p}_0$.*

Corollary 4 describes changes in the voter's threshold as a function of the strength of the IO's report s and his beliefs about the relative likelihood that each type sent an optimistic

message. The IO's dissemination of information to the public affects a voter's willingness to electorally support the incumbent, but is mediated by the voter's beliefs about the leader's initial message m. If, as intuition would suggest, the honest type is more likely to send $m = 1$ than the captured type, then a strong report from the IO, a high value of s, is a positive signal for the voter of the leader's type and therefore he sets a more lenient cutoff. But if the voter believes the captured type is more likely to send $m = 1$ than the honest type, a stronger report from the IO is actually a negative signal about the leader's honesty, and so the voter interprets the IO's report negatively, choosing a more stringent reelection threshold. Figure 1 illustrates these dynamics by plotting the voter's cutoff $\hat{x}(\hat{a}, \hat{p})$ as a function of s under three cases, when the voter believes the honest type to send $m = 1$ more than the captured type ($\hat{p}_1 > \hat{p}_0$), the opposite case ($\hat{p}_1 < \hat{p}_0$), and the case in which types send $m = 1$ with equal probability ($\hat{p}_1 = \hat{p}_0$).

Changing voter expectations as a function of the IO's report also determines the incumbent leader's first period effort. Similar to the model without the organization, the incumbent's reelection probability is equal to

$$\pi(a) = \Phi(\sqrt{\zeta}(a - \hat{x}(\hat{a}, \hat{p}))),$$

such that optimal efforts $a^*_\theta(\hat{p})$ balance

$$\underbrace{\beta}_{\text{marginal policy benefit}} + \underbrace{\sqrt{\zeta}\phi(\sqrt{\zeta}(a^*_\theta - \hat{x}(a^*, \hat{p})))\Psi}_{\text{signaling premium given IO}} = \underbrace{\beta\theta\lambda_M a^*_1 + \beta(1-\theta)(\lambda_M + \lambda_S)a^*_0}_{\text{marginal policy costs}}.$$

Results in Corollaries 1-3 all hold in this expanded model and the intuition for the optimal effort is the same. The difference is the threshold that the voter uses, reflecting updated beliefs based on conjectures about the leader's nationally determined contribution and the contents of the IO's report.

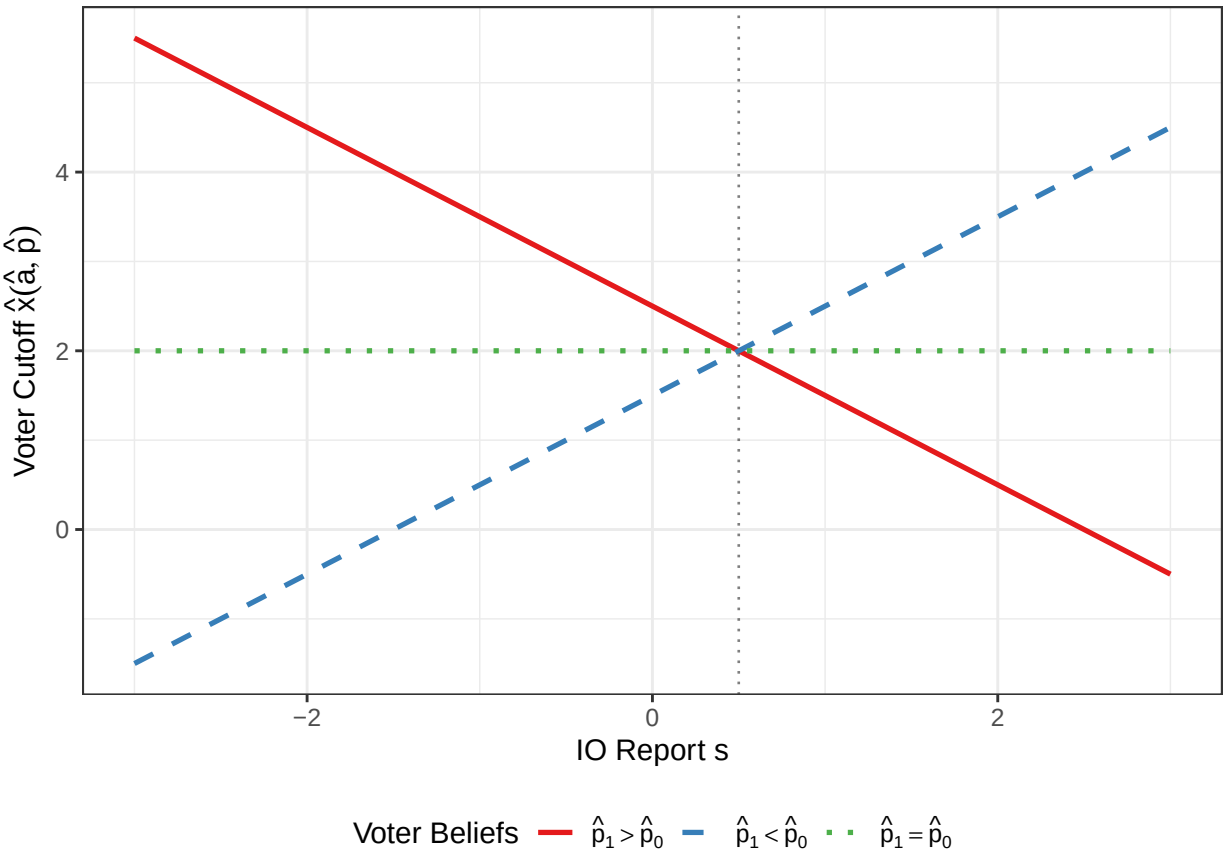

Figure 1: IO Reporting and Voter Beliefs of Leader Messages and the Voter's Cutoff
$\hat{a}_1 = 2.5$, $\hat{a}_0 = 1.5$, $\zeta = 1$, $\tau = 1$

To see how the IO affects effort choices, note that a strong report from the IO (strongly positive s) or a very weak report from the IO (strongly negative s) sends the voter's cutoff toward the extremes, $\hat{x}(\hat{a}, \hat{p}) \to \pm\infty$. If the voter is particularly lenient or stringent, then the incumbent leader knows the electoral outcome with near certainty, and the election is thus not conditioned on her effort choice. In these cases, the incumbent exerts effort at her ideal point. However, when s takes intermediate values, downstream effort is still a valuable means of signaling type to the voter. In fact, for intermediate s there are cases in which the leader exerts greater effort than in the model without the IO, as illustrated in Figure 2. If s is middling, meaning the voter gleans little information about the leader's type from the

IO's report, then the leader has additional incentives to ensure that the voter perceives her as honest, and thus exert greater effort.

Corollary 5. *Let $\tilde{a}_\theta^*$ be the equilibrium effort in the model without the organization. When comparing effort in the models with and without the organization:*

- *If $\hat{p}_1 = \hat{p}_0$, then effort is identical in both models, $a_\theta = \tilde{a}_\theta^*$.*
- *If $\hat{p}_1 \neq \hat{p}_0$, there exists $\underline{s}_\theta$ and $\overline{s}_\theta$ such that effort is larger in the model with the IO, $a_\theta^* \geq \tilde{a}_\theta^*$, when $s \in [\underline{s}_\theta,\ \overline{s}_\theta]$ and larger in the model without the IO otherwise.*

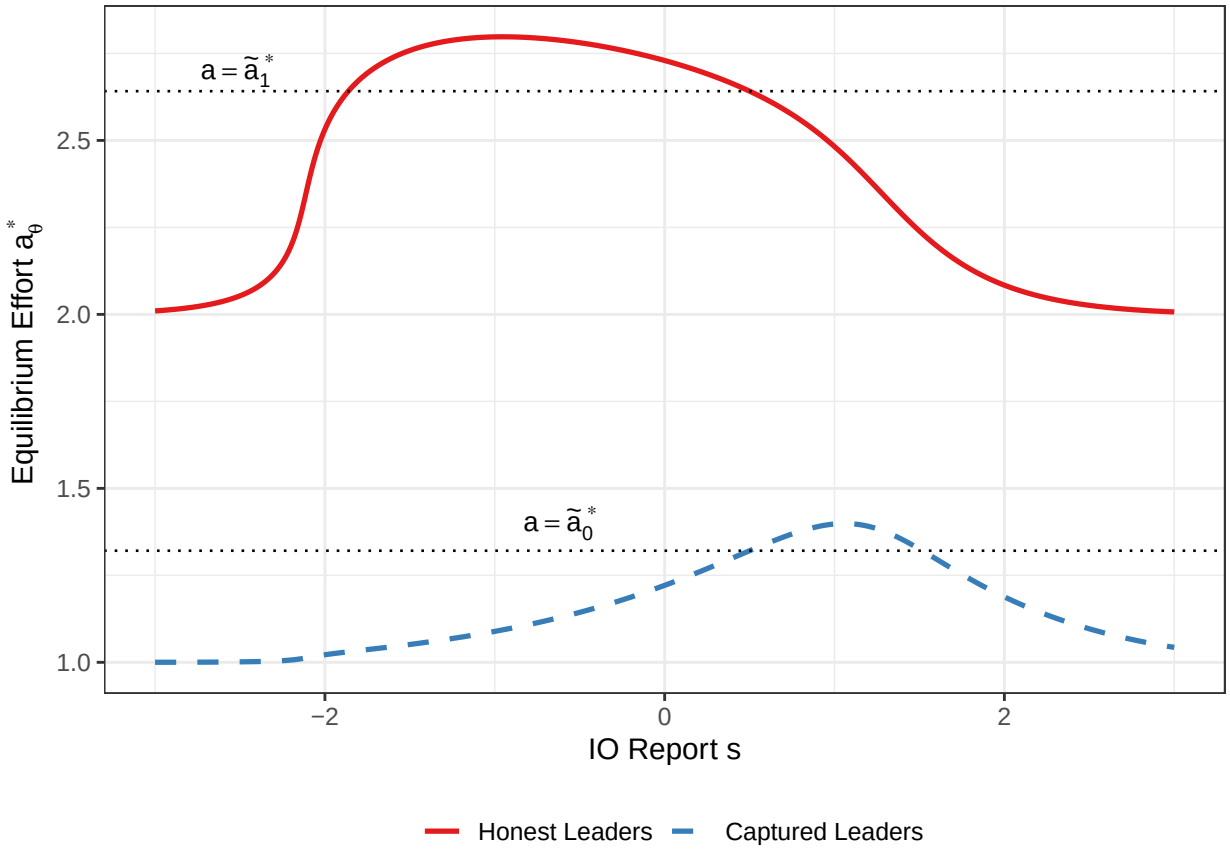

Figure 2: Leaders' Equilibrium Efforts and IO Reporting
$\lambda_M = \frac{1}{2}$, $\lambda_S = \frac{1}{2}$, $\beta = 1$, $\zeta = 1$, $\tau = 1$, $\Psi = 1$, $\hat{p}_1 = 1$, $\hat{p}_0 = 0$

In theory, this means that information disseminated by international organizations about leader pledges can inspire greater effort from leaders relative to a world without such an institution, conditional on the strength of the report that an IO sends. Unfortunately, this is not true on the equilibrium path: the next proposition details that leaders strategically engage with the IO, rendering its reports uninformative because they have incentives to obfuscate or signal jam when messaging their intentions globally.

Proposition 3. *In the unique equilibrium, both leaders send* $m = 1$ *with probability* $p^* \in (0, 1)$. *Equilibrium effort in the model with the IO is equivalent to that in the model without it.*

How do leaders determine which message to send to the IO? There are no direct costs to sending optimistic plans $m = 1$ or pessimistic plans $m = 0$, as these messages are merely plans or intentions about downstream effort. However, in equilibrium, the costs or implications of messages manifest in the voter's beliefs about the leader's type, and subsequently affect reelection probabilities. Therefore the leader's incentives to send message m rely on the maximization

$$\max_{m \in \{0,1\}} \int_{-\infty}^{\infty} \pi(a_\theta^*(s))\phi(\sqrt{\tau}(s-m))ds.$$

The IO attempts to act as a screening device—separating honest types from captured types by creating partially informative reports—but leaders anticipate the downstream effect of this information revelation. The key factor to consider is that leaders' messages must be mutually consistent with the beliefs that the voters has downstream about each type. To see why, consider the dynamics in Figure 1: if voter believes that honest leaders were more likely to send optimistic messages than captured leaders and the IO's report is poor, then the voter sets a stringent reelection cutoff (red solid line). This is not beneficial for honest leaders. By contrast, if the voter believes that captured leaders are more likely than honest leaders to report $m = 1$, then a very strong IO report engenders a stringent cutoff, which

harms captured leaders (blue dashed line). With an eye toward reelection, leaders seek to ensure that the voter uses the most favorable threshold to assess type, which on average occurs if the voter believes each type is equally likely to send $m = 1$.

The presence of the IO, interacted with downstream voter assessments of leader type, creates a chase-and-evade logic between honest leaders and captured leaders in equilibrium. Since talk is cheap, it is in the interest of captured leaders to mimic honest leaders, sending $m = 1$. Indeed, captured leaders may submit lofty proposals of effort to the IO without the intention of seeing such policies through (Melnick and Smith 2025). However, it is honest leaders, in an attempt to retain their *ex ante* electoral advantage, who wish to separate themselves from captured leaders. To do so, she would counterintuitively prefer to understate her willingness to cooperate as a means of distinguishing herself from a mimicking captured type vis-à-vis the voter. Hence, the leader's messaging is chosen to optimize electoral attractiveness: captured leaders prefer to pool and conceal their bias, while honest leaders want to separate. This is reminiscent of the classical incentives to misrepresent elsewhere in international politics, as honest leaders may find it beneficial to overstate her costs of exerting effort and report that they are captured (Fearon 1995; Schultz 1998; Powell 1999).

Given these competing temptations, the optimal strategy is for leaders to randomize between sending either message, and to do so with equal probability. Such a strategy is the only means through which both leaders can optimally satisfy their domestic constraints, meaning the downstream beliefs of the voter, which unfortunately renders the IO's report completely uninformative. This means that the equilibrium level of effort downstream is identical to that in the model without the IO.

Although nationally determined contributions are not informative in equilibrium, truthful revelation of of type occurs with probability p^* for honest leaders and probability $1 - p^*$ for captured leaders. Unfortunately, this truthfulness does not advance international cooperative goals of greater global public goods provision beyond domestic leaders' signaling incentives.

In short, the domestic political problem subordinates the international goal of incentivizing greater effort, as leaders strategically engage with international institutions to stifle any downstream informational effects of IO reporting on voter assessment of leader type.

Discussion and Conclusion

The equilibrium effects of information dissemination in pledge-and-review frameworks are striking. Information from IOs affects voter assessments of leader performance (Kelley and Simmons 2019), which subsequently influences how leaders engage with institutions in the first place. IOs that rely on informational strategies to induce compliance must be mindful not only of whether the information is credible or accurate, but also of how that information interacts with domestic political incentives. By way of conclusion, I discuss some of the domestic and international implications of the further proliferation of these institutions.

Given the maximal flexibility of pledge-and-review frameworks, it is clear *a priori* why policymakers prefer these institutional arrangements to more legalized forms of international cooperation in which binding targets are enforced pecuniarily (Dannenberg et al. 2023). In fact, the equilibrium analysis herein implies that leaders would be exactly indifferent over joining such an agreement and staying out. The large coalition, for which the Paris Agreement is primarily lauded (Calvo and Rubio 2013; Caparrós 2016; Harstad 2023*b*), is thus a symptom of leaders' strategic anticipation of the implications of pledges on domestic political evaluation. Hence, while these institutions might be an easy means of signing onto international collaborative efforts, they need not imply intentionality toward furthering goals of global public goods provision at all (Hollyer and Rosendorff 2011). Moreover, this finding contrasts with scholarship in which signing treaties is a signal of leaders' preferences for cooperation (e.g., Von Stein 2005; Simmons 2009). Through the interaction of institutional design and the desire for domestic political survival, leaders signal nothing with their

membership and engagement in such an organization.

The shortcomings of pledge-and-review stem not from strength but from structure. The precision of the IO's report τ plays no role in shaping equilibrium outcomes, which contrasts with extant studies emphasizing variation in institutional strength as a key determinant for how information dissemination shapes compliance (Chayes and Chayes 1993; Dai 2002; Hafner-Burton 2008). Leaders, cognizant of how international reports will influence their reelection odds, strategically obfuscate at the pledge-making stage in equilibrium.[8]

One path forward suggests that, rather than develop international cooperative structures on top of domestic policymaking efforts, international institutions could use their expertise to facilitate *domestic* transparency or accountability efforts on the implementation side (cf. Stavins and Stowe 2016, 43-47). In the model, there is an optimal level of transparency ζ that maximizes domestic contributions to global public goods and is commensurate with leaders' electoral incentives. Organizations could attempt to foster a closer connection between leaders' efforts to comply with international policies and the outputs of such efforts by, for example, increasing bureaucratic quality (Slough 2024). Beyond transparency, IOs could seek to shape the public's policy preferences (decreasing λ_M) through public-facing efforts that do not rely on information collection from national governments (see Gazmararian, Mildenberger and Tingley 2025), which would increase the social optimum for cooperative policies among the electorate. In this way, international organizations may facilitate cooperative goals without developing frameworks that create new opportunities for leaders to strategically obfuscate, but rather that are consistent with preexisting electoral incentives.

[8]Other work (e.g., Melnick and Smith 2025) shows how international shaming can influence downstream effort in the "review" stage, the shadow of which affects initial pledges.

References

Abbott, Kenneth W. and Duncan Snidal. 1998. "Why States Act through Formal International Organizations." *The Journal of Conflict Resolution* 42(1):3–32.

Abbott, Kenneth W., Philipp Genschel, Duncan Snidal and Bernhard Zangl, eds. 2015. *International Organizations as Orchestrators*. Cambridge: Cambridge University Press.

Acemoglu, Daron, Georgy Egorov and Konstantin Sonin. 2013. "A Political Theory of Populism." *The Quarterly Journal of Economics* 128(2):771–805.

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.

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

Awad, Emiel and Nicolás Riquelme. 2024. "International Cooperation, Information Transmission, and Delegation." *Quarterly Journal of Political Science* 19(2):217–242.

Bang, Guri, Jon Hovi and Tora Skodvin. 2016. "The Paris Agreement: Short-Term and Long-Term Effectiveness | Article | Politics and Governance.".

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

Bayer, Patrick and Lorenzo Crippa. 2025. "Government Influence in Information Production in International Organizations.".

Bechtel, Michael M. and Kenneth F. Scheve. 2013. "Mass support for global climate agree-

ments depends on institutional design." *Proceedings of the National Academy of Sciences* 110(34):13763–13768.

Bisbee, James H., James R. Hollyer, B. Peter Rosendorff and James Raymond Vreeland. 2019. "The Millennium Development Goals and Education: Accountability and Substitution in Global Assessment." *International Organization* 73(3):547–578.

Bolet, Diane, Fergus Green and Mikel González-Eguino. 2024. "How to Get Coal Country to Vote for Climate Policy: The Effect of a "Just Transition Agreement" on Spanish Election Results." *American Political Science Review* 118(3):1344 – 1359.

Bueno de Mesquita, Bruce, Alastair Smith, Randolph M. Siverson and James D. Morrow. 2003. *The Logic of Political Survival*. Cambridge: The MIT Press.

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.

Canes-Wrone, Brandice, Michael C. Herron and Kenneth W. Shotts. 2001. "Leadership and Pandering: A Theory of Executive Policymaking." *American Journal of Political Science* 45(3):532–550.

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.".

Chaudoin, Stephen. 2014. "Promises or Policies? An Experimental Analysis of International Agreements and Audience Reactions." *International Organization* 68(1):235–256.

Chayes, Abram and Antonia Handler Chayes. 1993. "On Compliance." *International Organization* 47(2):175–205.

Crippa, Lorenzo and Patrick Bayer. 2025. "The Politics of National Ties in Government Review of IO Reports.".

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

Dai, Xinyuan. 2005. "Why Comply? The Domestic Constituency Mechanism." *International Organization* 59(2):363–398.

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.

De Pryck, Kari and Mike Hulme, eds. 2022. *A Critical Assessment of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press.

Dechezleprêtre, Antoine, Adrien Fabre, Tobias Kruse, Bluebery Planterose, Ana Sanchez Chico and Stefanie Stantcheva. 2025. "Fighting Climate Change: International Attitudes toward Climate Policies." *American Economic Review* 115(4):1258–1300.

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.

Elliott, Christian, Steven Bernstein and Matthew Hoffmann. 2022. "Credibility dilemmas under the Paris agreement: explaining fossil fuel subsidy reform references in INDCs." *International Environmental Agreements: Politics, Law and Economics* 22(4):735–759.

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

Fang, Songying. 2008. "The Informational Role of International Institutions and Domestic Politics." *American Journal of Political Science* 52(2):304–321.

Fang, Songying and Randall W. Stone. 2012. "International Organizations as Policy Advisors." *International Organization* 66(4):537–569.

Fearon, James D. 1995. "Rationalist Explanations for War." *International Organization* 49(3):379–414.

Fearon, James D. 1999. Electoral Accountability and the Control of Politicians: Selecting Good Types versus Sanctioning Poor Performance. In *Democracy, Accountability, and Representation*, ed. Adam Przeworski, Bernard Manin and Susan C. Stokes. Cambridge Studies in the Theory of Democracy Cambridge: Cambridge University Press pp. 55–97.

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. 2025. "Public opinion foundations of the clean energy transition." *Environmental Politics* 0(0):1–23.

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

Grigorescu, Alexandru. 2003. "International Organizations and Government Transparency: Linking the International and Domestic Realms." *International Studies Quarterly* 47(4):643–667.

Grossman, Gene M. and Elhanan Helpman. 1994. "Protection for Sale." *The American Economic Review* 84(4):833–850.

Gruber, Lloyd. 2000. *Ruling the World: Power Politics and the Rise of Supranational Institutions*. Princeton: Princeton University Press.

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

Haftendorn, Edited by Helga, Robert O. Keohane and and Celeste A. Wallender, eds. 1999. *Imperfect Unions: Security Institutions Over Time and Space*. Oxford, New York: Oxford University Press.

Hale, Thomas and Charles Roger. 2014. "Orchestration and transnational climate governance." *The Review of International Organizations* 9(1):59–82.

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

Harstad, Bård. 2012. "Climate Contracts: A Game of Emissions, Investments, Negotiations, and Renegotiations." *The Review of Economic Studies* 79(4):1527–1557.

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.

Heddesheimer, Vincent, Hanno Hilbig and Erik Voeten. 2024. "The Green Transition and Political Polarization Along Occupational Lines.".

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.

Hollyer, James R., B. Peter Rosendorff and James Raymond Vreeland. 2015. "Transparency, Protest, and Autocratic Instability." *American Political Science Review* 109(4):764–784.

Hollyer, James R., B. Peter Rosendorff and James Raymond Vreeland. 2018. *Information, Democracy, and Autocracy: Economic Transparency and Political (In)Stability*. Cambridge: Cambridge University Press.

Hollyer, James R., B. Peter Rosendorff and James Raymond Vreeland. 2019*a*. "Transparency, Protest and Democratic Stability." *British Journal of Political Science* 49(4):1251–1277.

Hollyer, James R., B. Peter Rosendorff and James Raymond Vreeland. 2019*b*. "Why Do Autocrats Disclose? Economic Transparency and Inter-elite Politics in the Shadow of Mass Unrest." *Journal of Conflict Resolution* 63(6):1488–1516.

Hughes, Hannah. 2024. *The IPCC and the Politics of Writing Climate Change*. Cambridge: Cambridge University Press.

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

Kamil, Nila and Sylvia Karlsson-Vinkhuyzen. 2023. "Exploring the links between climate transparency and mitigation policy through a reflexive capacity lens: case studies of Brazil, China, India, Indonesia, and Mexico." *International Environmental Agreements: Politics, Law and Economics* 23(4):415–448.

Kelley, Judith and Beth Simmons. 2015. "Politics by Number: Indicators as Social Pressure in International Relations." *American Journal of Political Science* 59(1):55–70.

Kelley, Judith G. and Beth A. Simmons. 2019. "Introduction: The Power of Global Performance Indicators." *International Organization* 73(3):491–510.

Kennard, Amanda. 2020. "The Enemy of My Enemy: When Firms Support Climate Change Regulation." *International Organization* 74(2):187–221.

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

Keohane, Robert O. 1984. *After Hegemony: Cooperation and Discord in the World Political Economy*. Princeton: Princeton University Press.

Keohane, Robert O. and David G. Victor. 2011. "The Regime Complex for Climate Change." *Perspectives on Politics* 9(1):7–23.

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

Konrad, Susanne, Max van Deursen and Aarti Gupta. 2022. "Capacity building for climate transparency: neutral 'means of implementation' or generating political effects?" *Climate Policy* 22(5):557–575.

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

Lipscy, Phillip Y. 2017. *Renegotiating the World Order: Institutional Change in International Relations*. Cambridge: Cambridge University Press.

Mansfield, Edward D., Helen V. Milner and B. Peter Rosendorff. 2002. "Why Democracies Cooperate More: Electoral Control and International Trade Agreements." *International Organization* 56(3):477–513.

Maskin, Eric and Jean Tirole. 2004. "The Politician and the Judge: Accountability in Government." *American Economic Review* 94(4):1034–1054.

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

Melnick, Justin and Alastair Smith. 2025. "Shaming Paris: A Political Economy of Climate Commitments." *International Organization* 79(2):281–305.

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

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

Mitchell, Ronald B. 1998. "Sources of Transparency: Information Systems in International Regimes." *International Studies Quarterly* 42(1):109–130.

Mitchell, Ronald B. 2006. "Problem Structure, Institutional Design, and the Relative Effectiveness of International Environmental Agreements." *Global Environmental Politics* 6(3):72–89.

Nemet, Gregory F., Michael Jakob, Jan Christoph Steckel and Ottmar Edenhofer. 2017. "Addressing policy credibility problems for low-carbon investment." *Global Environmental Change* 42:47–57.

Paine, Jack and Scott A. Tyson. 2020. Uses and abuses of formal models in political science. In *The SAGE Handbook of Political Science*, ed. Dirk Berg-Schlosser, Bertrand Badie and Leonardo Morlino. Oxford: Oxford University Press pp. 188–202.

Pauw, W. P., Richard J. T. Klein, Kennedy Mbeva, Adis Dzebo, Davide Cassanmagnago and Anna Rudloff. 2018. "Beyond headline mitigation numbers: we need more transparent and comparable NDCs to achieve the Paris Agreement on climate change." *Climatic Change* 147(1):23–29.

Powell, Robert. 1999. *In the Shadow of Power: States and Strategies in International Politics*. Princeton: Princeton University Press.

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.

Raiser, Kilian, Ulrike Kornek, Christian Flachsland and William F Lamb. 2020. "Is the Paris Agreement effective? A systematic map of the evidence." *Environmental Research Letters* 15(8):083006.

Roberson, Brian. 2006. "The Colonel Blotto game." *Economic Theory* 29(1):1–24.

Rogelj, Joeri, Oliver Fricko, Malte Meinshausen, Volker Krey, Johanna J. J. Zilliacus and Keywan Riahi. 2017. "Understanding the origin of Paris Agreement emission uncertainties." *Nature Communications* 8:15748.

Rosendorff, B. Peter. 2005. "Stability and Rigidity: Politics and Design of the WTO's Dispute Settlement Procedure." *The American Political Science Review* 99(3):389–400.

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.

Schultz, Kenneth A. 1998. "Domestic Opposition and Signaling in International Crises." *The American Political Science Review* 92(4):829–844.

Simmons, Beth A. 2009. *Mobilizing for Human Rights: International Law in Domestic Politics*. Cambridge: Cambridge University Press.

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

Slough, Tara. 2024. "Bureaucratic Quality and Electoral Accountability." *American Political Science Review* 118(4):1931–1950.

Stavins, Robert N. and Robert C. Stowe, eds. 2016. *The Paris Agreement and Beyond: International Climate Change Policy Post-2020*. Cambridge: Harvard Project on Climate Agreements.

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.

Stone, Randall W. 2011. *Controlling Institutions: International Organizations and the Global Economy*. Cambridge: Cambridge University Press.

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 and Vegard Wiborg. 2024. "Commitment ambiguity and ambition in climate pledges." *The Review of International Organizations* .

UNFCCC. 2015. "Paris Agreement.".

van Deursen, Max and Aarti Gupta. 2025. "Is Enhanced Transparency the "Backbone" of the Paris Agreement? A Critical Assessment." *Transnational Environmental Law* 14(1):41–68.

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.

Voeten, Erik. 2025. "The Energy Transition and Support for the Radical Right: Evidence from the Netherlands." *Comparative Political Studies* 58(2):394–428.

Von Stein, Jana. 2005. "Do Treaties Constrain or Screen? Selection Bias and Treaty Compliance." *The American Political Science Review* 99(4):611–622.

Appendix

Proof of Proposition 1. Proceed by backward induction. In the second period, a leader of type θ maximizes

$$\max_{a_2(\theta)} \beta\Big(a_2(\theta) - \theta\frac{\lambda_M}{2}a_2(\theta)^2 - (1-\theta)\frac{\lambda_M+\lambda_S}{2}a_2(\theta)^2\Big).$$

The first-order condition is

$$\beta\Big(1 - \theta\lambda_M a_2(\theta) - (1-\theta)(\lambda_M+\lambda_S)a_2(\theta) = 0,$$

which has solution $a_2^*(1) = \frac{1}{\lambda_M}$ and $a_2^*(0) = \frac{1}{\lambda_M+\lambda_S}$. These choices are a maximum because the leader's utility function is globally concave, as the second-order condition is

$$-\theta\lambda_M - (1-\theta)(\lambda_M+\lambda_S) < 0.$$

Since $a_2^*(1) > a_2^*(0)$ and in particular $a_2^*(1)$ is the median voter's ideal point, the median voter wants to retain the incumbent leader when his posterior belief about the leader's honesty is greater than the prior. Moreover, since x_1 is FOSD-increasing in a_1, higher signals are on average more likely to signal honesty. Therefore the voter prefers to retain the incumbent whenever the signal x_1 is greater than some threshold $\hat{x}$. Let $\mu(x) = P(\theta = 1|x_1 = x)$ be the voter's posterior belief that the incumbent is honest given the realized policy outcome $x_1 = x$. As effort is unobserved, let the voter have conjecture about the incumbent's effort choice, $\hat{a}_1(\theta)$. Formally, posterior beliefs can be expressed as

$$\mu(x) = \frac{\gamma\phi(\sqrt{\zeta}(x - \hat{a}_1(1)))}{\gamma\phi(\sqrt{\zeta}(x - \hat{a}_1(1))) + (1-\gamma)\phi(\sqrt{\zeta}(x - \hat{a}_1(0)))}.$$

The voter retains the incumbent iff $\mu(x) \geq \gamma$, which is equivalent to

$$x \geq \frac{\hat{a}_1 + \hat{a}_0}{2}.$$

Given $x_1 = a_1 + \varepsilon_1$, the incumbent leader survives iff $a_1 + \varepsilon_1 \geq \frac{\hat{a}_1 + \hat{a}_0}{2}$. Since $\varepsilon_1 \sim N(0, \frac{1}{\zeta})$, the incumbent's reelection probability is equal to

$$\pi(a_1) = \Phi(\sqrt{\zeta}(a_1 - \frac{\hat{a}_1(1) + \hat{a}_1(0)}{2})).$$

In the first period, the leader of type θ maximizes

$$\max_{a_1(\theta)} \ \beta\Big(a_1(\theta) - \theta\frac{\lambda_M}{2}a_1(\theta)^2 - (1-\theta)\frac{\lambda_M + \lambda_S}{2}a_1(\theta)^2\Big) + \pi(a_1(\theta))\Psi,$$

which leads to the first-order condition

$$\beta - \theta\lambda_M a_1(\theta) - (1-\theta)(\lambda_M + \lambda_S)a_1(\theta) + \sqrt{\zeta}\phi(\sqrt{\zeta}(a_1 - \frac{\hat{a}_1(1) + \hat{a}_1(0)}{2}))\Psi = 0.$$

Since beliefs are correct in equilibrium, $a_1(\theta) = \hat{a}_1(\theta) = a_1^*(\theta)$, this simplifies to

$$\beta - \theta\lambda_M a_1(\theta) - (1-\theta)(\lambda_M + \lambda_S)a_1(\theta) + \sqrt{\zeta}\phi(\sqrt{\zeta}(\frac{a_1^*(1) + a_1^*(0)}{2}))\Psi = 0.$$

Substituting in $\theta = 1$ and $\theta = 0$ yields the two equations in the proposition.

To show that this solution is a maximum, we ensure that the leader's utility is concave. The second-order condition is

$$-\theta\lambda_M - (1-\theta)(\lambda_M + \lambda_S) + \zeta^{3/2}(a_1 - \frac{\hat{a}_1(1) + \hat{a}_1(0)}{2})\phi(\sqrt{\zeta}(a_1 - \frac{\hat{a}_1(1) + \hat{a}_1(0)}{2}))\Psi.$$

Let $\eta = \sqrt{\zeta}(a_1 - \frac{\hat{a}_1(1)+\hat{a}_1(0)}{2})$ so the second-order condition can be rewritten as

$$-\lambda_M - (1-\theta)\lambda_S + \zeta\eta\phi(\eta)\Psi.$$

The standard normal density tends to zero faster than any polynomial so $\eta\phi(\eta)$ is zero at $\eta = 0$ and approaches zero as $\eta \to \pm\infty$. The derivative of $\eta\phi(\eta)$ is $\phi(\eta) - \eta^2\phi(\eta)$ with critical points at $\eta = \pm 1$. Note that if $\eta = -1$ then the problem is globally concave. Hence the relevant constraint is at $\eta = 1$, where $\eta\phi(\eta) = \frac{1}{\sqrt{2\pi e}}$. Hence the leader's utility is concave iff

$$-\lambda_M - (1-\theta)\lambda_S + \frac{\zeta}{\sqrt{2\pi e}}\Psi < 0,$$

or $\zeta < \frac{\lambda_M + (1-\theta)\lambda_S\sqrt{2\pi e}}{\Psi}$. Hence a sufficient condition for both leaders to have concave utility functions is $\zeta < \frac{\lambda_M\sqrt{2\pi e}}{\Psi}$.

Furthermore, this equilibrium is unique because pooling cannot be an equilibrium. By way of contradiction, suppose that the voter believed $\hat{a}_1(\theta) = \hat{a}$ for any θ. Then $\mu(x) = \gamma$ for any x, and the voter is indifferent between retaining and replacing the incumbent. Hence, depending on how ties are broken, the incumbent's reelection probability is either zero or 1. This means that the incumbent leader's maximization problem is equivalent to

$$\max_{a_1(\theta)} \ \beta\Big(a_1(\theta) - \theta\frac{\lambda_M}{2}a_1(\theta)^2 - (1-\theta)\frac{\lambda_M + \lambda_S}{2}a_1(\theta)^2\Big),$$

the solution to which is $a_1(1) = \frac{1}{\lambda_M}$ and $a_1(0) = \frac{1}{\lambda_M + \lambda_S}$ such that $a_1(1) \neq a_1(0)$. □

Proof of Corollary 1. From Proposition 1, leader's effort choices satisfy

$$\beta + \sqrt{\zeta}\phi(\sqrt{\zeta}(\frac{a_1^*(1) - a_1^*(0)}{2}))\Psi = \beta\lambda_M a_1^*(1).$$
$$\beta + \sqrt{\zeta}\phi(\sqrt{\zeta}(\frac{a_1^*(0) - a_1^*(1)}{2}))\Psi = \beta(\lambda_M + \lambda_S)a_1^*(0).$$

The LHS of these equations are the same, which implies that $\beta\lambda_M a_1^*(1) = \beta(\lambda_M + \lambda_S)a_1^*(0)$, or $a_1^*(1) = \frac{\lambda_M+\lambda_S}{\lambda_M}a_1^*(0)$ so $a_1^*(1) > a_1^*(0)$.

To see that $a_1^*(1) > \frac{1}{\lambda_M}$, substitute $a_1 = \frac{1}{\lambda_M}$ into the honest leader's first-order condition to get

$$\sqrt{\zeta}\phi(\sqrt{\zeta}(\frac{1/\lambda_M - a_1^*(0)}{2}))\Psi > 0,$$

which holds for any $a_1^*(0)$. Since this first-order condition is positive, we must have $a_1^*(1) > \frac{1}{\lambda_M}$. Similarly, substituting $a_1 = \frac{1}{\lambda_M}$ into the captured leader's first-order condition yields

$$\beta(1 - \frac{\lambda_M + \lambda_S}{\lambda_M}) + \sqrt{\zeta}\phi(\sqrt{\zeta}(\frac{1/\lambda_M - a_1^*(1)}{2}))\Psi.$$

This expression can be either positive or negative. Note that the standard normal density takes a maximum value of $\frac{1}{\sqrt{2\pi}}$, and so a sufficient condition for the captured leader's equilibrium effort to be larger than $\frac{1}{\lambda_M}$ is

$$\beta(1 - \frac{\lambda_M + \lambda_S}{\lambda_M}) + \sqrt{\frac{\zeta}{2\pi}}\Psi > 0,$$

which occurs whenever $\lambda_S < \lambda_M + \sqrt{\frac{\zeta}{2\pi}}\frac{\Psi}{\beta}$. □

Proof of Corollary 2. Follows from x_1 FOSD-increasing in a_1 and $a_1^*(1) > a_1^*(0)$. □

Proof of Corollary 3. Define the Jacobian for type θ as

$$\mathbf{J}_\theta = \begin{bmatrix} \frac{\partial^2 v_\theta}{\partial a_\theta^2} & \frac{\partial^2 v_\theta}{\partial a_\theta \partial a_{\theta'}} \\ \frac{\partial^2 v_{\theta'}}{\partial a_\theta \partial a_{\theta'}} & \frac{\partial^2 v_{\theta'}}{\partial a_{\theta'}^2} \end{bmatrix}.$$

Observe that $\frac{\partial^2 v_1}{\partial a_1^2} < 0$ and $\frac{\partial^2 v_0}{\partial a_0^2} < 0$ at the equilibrium (a_1^*, a_0^*) because they are maxima. Further observe that $\frac{\partial^2 v_1}{\partial a_1 \partial a_0} > 0$ and $\frac{\partial^2 v_0}{\partial a_1 \partial a_0} < 0$, hence $|\mathbf{J}_\theta| > 0$. Given this structure, the direct and indirect effects have the same sign; without loss of generality I simply consider

the direct effects.

By monotone comparative statics, taking the cross-partial of the leader's utility with respect to parameters yields

$$\begin{aligned}
\frac{\partial^2 v_L(a_t;\theta)}{\partial a_1 \partial \Psi} &= \sqrt{\zeta}\phi(\sqrt{\zeta}(a_1 - \frac{\hat{a}_1(1)+\hat{a}_1(0)}{2})) \geq 0. \\
\frac{\partial^2 v_L(a_t;\theta)}{\partial a_1 \partial \lambda_M} &= -a_1 \leq 0. \\
\frac{\partial^2 v_L(a_t;\theta)}{\partial a_1 \partial \lambda_S} &= -(1-\theta)a_1 \leq 0. \\
\frac{\partial^2 v_L(a_t;\theta)}{\partial a_1 \partial \zeta} &= \frac{\Psi}{2\sqrt{\zeta}}\phi(\sqrt{\zeta}(a_1 - \frac{\hat{a}_1(1)+\hat{a}_1(0)}{2}))\Big(1-\zeta(a_1 - \frac{\hat{a}_1(1)+\hat{a}_1(0)}{2})^2\Big).
\end{aligned}$$

These inequalities imply that effort a_θ^* is increasing in Ψ, decreasing in λ_M, and a_0^* is decreasing in λ_S. Furthermore, while the direct effect $\frac{\partial a_1^*}{\partial \lambda_S} = 0$, the indirect effect from a_0^* is such that a_1^* is decreasing in λ_S as well. Also observe that $\Big(1-\zeta(a_1 - \frac{\hat{a}_1(1)+\hat{a}_1(0)}{2})^2\Big) > 0$ is positive as $\zeta \to 0$ and decreasing in ζ so that the effect is inverse U-shaped. □

Proof of Proposition 2. Proof is analogous to that of Proposition 1. The only difference is the derivation of the voter's policy cutoff, which is a function of conjectures about the leader's effort $\hat{a}_\theta$ as well as conjectures about the messages sent to the IO $\hat{p}_\theta$.

Denote $\mu(x,s)$ as the voter's posterior belief about the leader's type having observed IO report s and signal x of the leader's effort. Since the leader's true message m and true effort a are unobserved, the voter needs to have conjectures. Let $\hat{a}_\theta$ be the voter's conjecture about leader-type θ's effort, and let $\hat{p}_\theta = P(m=1|\theta)$ be the voter's conjecture about the probability that leader-type θ sent message $m=1$ to the IO. Then $\hat{m}_\theta = \hat{p}_\theta\phi(\sqrt{\tau}(s-1)) + (1-\hat{p}_\theta)\phi(\sqrt{\tau}s)$ is the total probability that that IO's report is realized as the value s given voter's conjectures. Then $\mu(x,s)$ can be expressed as

$$\mu(x,s) = \frac{\gamma\phi(\sqrt{\zeta}(x-\hat{a}_1))\hat{m}_1}{\gamma\phi(\sqrt{\zeta}(x-\hat{a}_1))\hat{m}_1 + (1-\gamma)\phi(\sqrt{\zeta}(x-\hat{a}_0))\hat{m}_0},$$

such that it is optimal to retain the incumbent leader whenever

$$x \geq \frac{\hat{a}_1 + \hat{a}_0}{2} + \frac{\log(\frac{\hat{m}_0}{\hat{m}_1})}{\zeta(\hat{a}_1 - \hat{a}_0)} \equiv \hat{x}(\hat{a}, \hat{p}).$$

Given this cutoff the rest of the proof is identical with an identical characterization of the optimal first period effort. □

Proof of Corollary 4. Follows from the fact that the IO's report s has the MLRP in $m(\theta)$. □

Proof of Corollary 5. Recall that the voter's cutoff is defined as

$$\hat{x}(\hat{a}, \hat{p}) = \frac{\hat{a}_1 + \hat{a}_0}{2} + \frac{\log(\frac{\hat{m}_0}{\hat{m}_1})}{\zeta(\hat{a}_1 - \hat{a}_0)}.$$

It is immediate that whenever $\hat{p}_1 = \hat{p}_0$ then $\hat{x}(\hat{a}, \hat{p}) = \frac{\hat{a}_1 + \hat{a}_0}{2}$, as in the model without the IO. Optimal effort is thus identical to that characterized in Proposition 1.

Suppose $\hat{m}_1 \neq \hat{m}_0$. The first-order condition for leader-type θ's effort is

$$\beta - \beta\theta\lambda_M a - \beta(1-\theta)(\lambda_M + \lambda_S)a + \sqrt{\zeta}\phi(\sqrt{\zeta}(a - \frac{\hat{a}_1 + \hat{a}_0}{2} - \frac{\log(\frac{\hat{m}_0}{\hat{m}_1})}{\zeta(\hat{a}_1 - \hat{a}_0)}))\Psi = 0.$$

Since the normal density is log-concave, it is single peaked. Hence $\phi(\sqrt{\zeta}(a - \hat{x}(\hat{a}, \hat{p})))$ is single peaked in s such that there is a s^{max} where $\frac{d}{ds}\phi(\sqrt{\zeta}(a - \hat{x}(\hat{a}, \hat{p}))) > 0$ for $s < s^{max}$ and $\frac{d}{ds}\phi(\sqrt{\zeta}(a - \hat{x}(\hat{a}, \hat{p}))) < 0$ for $s > s^{max}$. As such optimal effort is single peaked in s, $\frac{da^*_\theta}{ds}$ is nonmonotonic in s. Moreover, observe that $\lim_{s\to-\infty}\phi(\sqrt{\zeta}(a - \hat{x}(\hat{a}, \hat{p}))) = 0$ and $\lim_{s\to\infty}\phi(\sqrt{\zeta}(a - \hat{x}(\hat{a}, \hat{p}))) = 0$ such that as $s \to \pm\infty$, $a^*_\theta \to \frac{1}{\lambda_M + (1-\theta)\lambda_S}$.

Denote leader-type θ's optimal effort in the model without the IO as $\tilde{a}_\theta$. Therefore since a^*_θ is continuous in s and $\tilde{a}_\theta > \frac{1}{\lambda_M + (1-\theta)\lambda_S}$ there exists $\underline{s}_\theta$ such that $a^*_\theta = \tilde{a}_\theta$ when $\frac{da^*_\theta}{ds} > 0$ and $\overline{s}_\theta$ such that $a^*_\theta = \tilde{a}_\theta$ when $\frac{da^*_\theta}{ds} < 0$. □

Lemma 1. *If $\hat{p}_1 \neq \hat{p}_0$, the voter's threshold $\hat{x}(\hat{a}, \hat{p})$:*

- *increases in* $\hat{p}_1$ *if* $s < \frac{1}{2}$ *and decreases in* $\hat{p}_1$ *if* $s > \frac{1}{2}$*;*
- *decreases in* $\hat{p}_0$ *if* $s < \frac{1}{2}$ *and increases in* $\hat{p}_0$ *if* $s > \frac{1}{2}$.

Proof of Lemma 1. The voter's threshold is

$$\hat{x}(\hat{a}, \hat{p}) = \frac{\hat{a}_1 + \hat{a}_0}{2} + \frac{\log(\frac{\hat{m}_0}{\hat{m}_1})}{\zeta(\hat{a}_1 - \hat{a}_0)},$$

where $\hat{m}_\theta = \hat{p}_\theta \phi(\sqrt{\tau}(s-1)) + (1-\hat{p}_\theta)\phi(\sqrt{\tau}s)$. Observe that

$$\frac{\partial \hat{m}_\theta}{\partial \hat{p}_\theta} = \phi(\sqrt{\tau}(s-1)) - \phi(\sqrt{\tau}s),$$

which is negative if $s < \frac{1}{2}$ and positive if $s > \frac{1}{2}$.

Differentiating with respect to $\hat{p}_1$ yields

$$\frac{\partial \hat{x}(\hat{a}, \hat{p})}{\partial \hat{p}_1} = -\frac{1}{\zeta(\hat{a}_1 - \hat{a}_0)\hat{m}_1} \frac{\partial \hat{m}_1}{\partial \hat{p}_1},$$

such that $\frac{\partial \hat{x}(\hat{a}, \hat{p})}{\partial \hat{p}_1} > 0$ if $s < \frac{1}{2}$ and $\frac{\partial \hat{x}(\hat{a}, \hat{p})}{\partial \hat{p}_1} < 0$ if $s > \frac{1}{2}$.

Similarly, differentiating with respect to $\hat{p}_0$ yields

$$\frac{\partial \hat{x}(\hat{a}, \hat{p})}{\partial \hat{p}_0} = \frac{1}{\zeta(\hat{a}_1 - \hat{a}_0)\hat{m}_0} \frac{\partial \hat{m}_0}{\partial \hat{p}_0},$$

such that $\frac{\partial \hat{x}(\hat{a}, \hat{p})}{\partial \hat{p}_0} < 0$ if $s < \frac{1}{2}$ and $\frac{\partial \hat{x}(\hat{a}, \hat{p})}{\partial \hat{p}_0} > 0$ if $s > \frac{1}{2}$. □

Proof of Proposition 3. The leader maximizes

$$\max_{m \in \{0,1\}} \int_{-\infty}^{\infty} \Big[\beta\Big(a_\theta^* - \frac{\lambda_M + (1-\theta)\lambda_S}{2} a_\theta^{*^2}\Big) + \pi(a_\theta^*(s))\Psi\Big]\phi(\sqrt{\tau}(s-m))ds,$$

therefore choosing $m = 1$ over $m = 0$ whenever

$$\int_{-\infty}^{\infty} \Big[\beta\Big(a_\theta^* - \frac{\lambda_M + (1-\theta)\lambda_S}{2} {a_\theta^*}^2\Big) + \pi(a_\theta^*(s))\Psi\Big]\phi(\sqrt{\tau}(s-1))ds \geq \int_{-\infty}^{\infty} \Big[\beta\Big(a_\theta^* - \frac{\lambda_M + (1-\theta)\lambda_S}{2} {a_\theta^*}^2\Big) + \pi(a_\theta^*(s))\Psi\Big]\phi(\sqrt{\tau}s)ds,$$

which simplifies to

$$\int_{-\infty}^{\infty} \pi(a_\theta^*(s))\Big(\phi(\sqrt{\tau}(s-1)) - \phi(\sqrt{\tau}s)\Big)ds \geq 0.$$

Define $\Delta_\theta(\hat{p}_1, \hat{p}_0) = \int_{-\infty}^{\infty} \pi(a_\theta^*(s))\Big(\phi(\sqrt{\tau}(s-1)) - \phi(\sqrt{\tau}s)\Big)ds$ as the leader's difference in expected reelection probability from sending message $m = 1$ versus $m = 0$ when she is of type θ. If $\hat{p}_1 = \hat{p}_0$, then $\hat{x}(a^*, \hat{p}) = \frac{a_1^* + a_0^*}{2}$, and $\pi(a_\theta^*; s)$ is constant in s so $\Delta_\theta(\hat{p}_1, \hat{p}_0)$ is the difference of two densities integrated over their entire support, thus $\Delta_\theta(\hat{p}_1, \hat{p}_0) = 0$. If $\Delta_\theta(\hat{p}_1, \hat{p}_0) = 0$, it must be because $\hat{p}_1 = \hat{p}_0$. Observe that $\pi(a^*; s) = 0$ only if $s \to \pm\infty$, so for any finite s $\pi(a^*; s) > 0$. Moreover we are integrating over the entire space of s so it must be that $\pi(a^*; s)$ is constant in s and $\int_{-\infty}^{\infty} \Big(\phi(\sqrt{\tau}(s-1)) - \phi(\sqrt{\tau}s)\Big)ds = 0$, which occurs when $\hat{p}_1 = \hat{p}_0$. Hence $\Delta_\theta(\hat{p}_1, \hat{p}_0) = 0$ iff $\hat{p}_1 = \hat{p}_0$.

Now we show that $\hat{p}_1 = \hat{p}_0$ must occur at an interior $p^* \in (0, 1)$. For the honest type,

$$\frac{\partial \Delta_1(\hat{p}_1, \hat{p}_0)}{\partial \hat{p}_1} = \int_{-\infty}^{\infty} \sqrt{\zeta}\phi(\sqrt{\zeta}(a_1^* - \hat{x}(a^*, \hat{p}))) \frac{1}{\zeta(a_1^* - a_0^*)\hat{m}_1} \Big(\phi(\sqrt{\tau}(s-1)) - \phi(\sqrt{\tau}s)\Big)^2 ds > 0,$$

so increasing the voter's belief that the honest type sends $m = 1$ increases the return from playing $m = 1$ versus $m = 0$. For the captured type,

$$\frac{\partial \Delta_0(\hat{p}_1, \hat{p}_0)}{\partial \hat{p}_0} = \int_{-\infty}^{\infty} -\sqrt{\zeta}\phi(\sqrt{\zeta}(a_0^* - \hat{x}(a^*, \hat{p}))) \frac{1}{\zeta(a_1^* - a_0^*)\hat{m}_0} \Big(\phi(\sqrt{\tau}(s-1)) - \phi(\sqrt{\tau}s)\Big)^2 ds < 0.$$

From this we know that $\Delta_1(\hat{p}_1, \hat{p}_0) < 0$ if $\hat{p}_1 < \hat{p}_0$ and $\Delta_1(\hat{p}_1, \hat{p}_0) > 0$ if $\hat{p}_1 > \hat{p}_0$. Furthermore, $\Delta_0(\hat{p}_1, \hat{p}_0) > 0$ if $\hat{p}_1 > \hat{p}_0$ and $\Delta_1(\hat{p}_1, \hat{p}_0) < 0$ if $\hat{p}_1 < \hat{p}_0$. To see that $\hat{p}_1 = \hat{p}_0 = 1$

or $\hat{p}_1 = \hat{p}_0 = 0$ cannot be an equilibrium, observe that $\Delta_1(\hat{p}_1, 1) < 0$ for any $\hat{p}_1$, meaning the honest type would deviate to $m = 0$. Similarly, $\Delta_1(\hat{p}_1, 0) > 0$ for any $\hat{p}_0$, meaning the captured type would deviate to $m = 1$. Thus the only equilibrium is $p_1^* = p_0^* = p \in (0, 1)$. $\square$

A basic extension of the model allows for the inclusion of a valence shock. In addition to the utility specified in the main text, consider a case where the voter also has a predisposed bias toward the incumbent leader, which represents the value of the incumbent on all other dimensions besides the implementation of the public good. Denote bias as y_i, where $y_i \sim U[-\alpha,\ \alpha]$. The value of this bias is realized right before the voter makes his choice to retain the leader or not.

Proposition A.1. *The inclusion of the valence shock y_i produces qualitatively equivalent results to the domestic politics model in the main text.*

Proof of Proposition A.1. Second period behavior is unaffected. The voter i adopts a decision rule in which he retains the leader if and only if

$$P(\theta = 1|x) + y_i \geq \gamma.$$

Posterior beliefs $P(\theta = 1|x)$ are equivalent to those derived above in the proof of Proposition 1. Conditional on some value of his bias y_i, the voter is thus exactly indifferent between retaining the incumbent leader and replacing her when

$$x(y_i) = \frac{\hat{a}_1 + \hat{a}_0}{2} + \frac{\log\left(\frac{(1-\gamma)(\gamma - y_i)}{\gamma(1-\gamma+y_i)}\right)}{\zeta(\hat{a}_1 - \hat{a}_0)}.$$

The likelihood ratio $\frac{\phi(\sqrt{\zeta}(x_i - \hat{a}_1))}{\phi(\sqrt{\zeta}(x_i - \hat{a}_0)}$ is increasing in the signal x_i. Therefore, the voter in country i retains his leader if and only if $x_i \geq \hat{x}$. Also note that if $y_i > \gamma$ then $x(y_i) \to -\infty$ and if $y_i < -1 + \gamma$ then $x(y_i) \to \infty$. The threshold $\hat{x}$ that the voter uses to reelect the incumbent

is

$$\hat{x}(y_i) = \begin{cases} \infty & y_i < -1+\gamma \\ \frac{\hat{a}_1+\hat{a}_0}{2} + \frac{\log\left(\frac{(1-\gamma)(\gamma-y_i)}{\gamma(1-\gamma+y_i)}\right)}{\zeta(\hat{a}_1-\hat{a}_0)} & -1+\gamma < y_i < \gamma \\ -\infty & y_i > \gamma. \end{cases}$$

Clearly, this means that if $y_i > \gamma$ the leader is retained with probability 1 and if $y_i < -1+\gamma$ the leader is retained with probability zero. This means that the leader's effort can only affect the outcome of the election if bias is moderate, or when $-1+\gamma < y_i < \gamma$. Therefore, the probability of reelection can be decomposed into two terms. If $y_i > \gamma$, the leader survives with probability 1, which occurs with $P(y_i > \gamma) = \frac{\alpha-\gamma}{2\alpha}$. Second, if $-1+\gamma < y_i < \gamma$, the leader survives with probability $\Phi(\sqrt{\zeta}(a_i - \hat{x}(y_i)))$. Therefore, the total probability of survival in office is

$$\frac{1}{2\alpha}\int_{-1+\gamma}^{\gamma} \Phi(\sqrt{\zeta}(a_i - \hat{x}(y_i)))\ dy + \frac{\alpha-\gamma}{2\alpha}.$$

Leader i maximizes the following expected utility:

$$EU_i(a;\theta_i) = u(a;\theta_i) + \Big[\int_{-1+\gamma}^{\gamma} \Phi(\sqrt{\zeta}(a_i - \hat{x}(y_i)))\ dy + \alpha - \gamma\Big]\frac{\Psi}{2\alpha},$$

where $u(a;\theta_i)$ is the leader's policy utility as defined in the main text.

For type θ_i, the first-order condition is

$$\frac{\partial u(a;\theta_i)}{\partial a_i} + \frac{\sqrt{\zeta}\Psi}{2\alpha}\int_{-1+\gamma}^{\gamma} \phi\Big(\sqrt{\zeta}(a_i - \frac{\hat{a}_1+\hat{a}_0}{2} - \frac{\log\left(\frac{(1-\gamma)(\gamma-y_i)}{\gamma(1-\gamma+y_i)}\right)}{\zeta(\hat{a}_1-\hat{a}_0)})\Big)\ dy = 0.$$

Equilibrium requires that voters' conjectures are correct, so this simplifies to

$$\frac{\partial u(a;\theta_i)}{\partial a_i} + \frac{\sqrt{\zeta}\Psi}{2\alpha}\int_{-1+\gamma}^{\gamma} \phi\Big(\sqrt{\zeta}(\frac{a_1^*+a_0^*}{2} - \frac{\log\left(\frac{(1-\gamma)(\gamma-y_i)}{\gamma(1-\gamma+y_i)}\right)}{\beta(a_1^*-a_0^*)})\Big)\ dy = 0.$$

Because leaders/countries are symmetric, there are 2 equations in 2 unknowns. Solving these equations yield optimal effort levels (a_1^*, a_0^*). To confirm that the equilibrium policy choices are a maximum, I take the second-order condition. Define $\eta(a_i, y_i) = \sqrt{\zeta}(a_i - \frac{\hat{a}_1+\hat{a}_0}{2} - \frac{\log\left(\frac{(1-\gamma)(\gamma-y_i)}{\gamma(1-\gamma+y_i)}\right)}{\zeta(\hat{a}_1-\hat{a}_0)})$. Using the fact that $\frac{d}{da}\phi(\eta) = -\eta\phi(\eta)\frac{\partial\eta}{\partial a}$, the second-order condition is

$$-\frac{\partial^2 u(a;\theta)}{\partial a_i^2} - \frac{\zeta\Psi}{2\alpha}\int_{-1+\gamma}^{\gamma} \eta(a_i, y_i)\phi(\eta(a_i, y_i))\, dy.$$

Note that $\eta(a_1^*, y_i) = \frac{a_1^*-a_0^*}{2} - \frac{\log\left(\frac{(1-\gamma)(\gamma-y_i)}{\gamma(1-\gamma+y_i)}\right)}{\zeta(\hat{a}_1-\hat{a}_0)}) > 0$. Therefore the function inside the integral in the second-order condition for the honest type is always positive, meaning the second-order condition $\frac{\partial^2 u(a;\theta)}{\partial a_i^2} - \frac{\zeta\Psi}{2\alpha}\int_{-1+\gamma}^{\gamma}\eta(a_1^*, y_i)\phi(\eta(a_1^*, y_i))\, dy < 0$ for the honest type.

Now consider the second-order condition for the captured type. Note that $\eta(a_0^*, y_i) = \frac{a_0^*-a_1^*}{2} - \frac{\log\left(\frac{(1-\gamma)(\gamma-y_i)}{\gamma(1-\gamma+y_i)}\right)}{\zeta(\hat{a}_1-\hat{a}_0)})$ need not be positive. A sufficient condition to show that the equilibrium effort a_0^* is a maximum is to find a lower bound on the integral. Differentiating $\eta(a_0^*, y_i)\phi(\eta(a_0^*, y_i))$ with respect to y_i yields the critical points $y_i = \frac{\gamma-1}{\frac{1}{\gamma e^{\frac{1}{2}\zeta(a_0^*-a_1^*)^2+\sqrt{\zeta}(a_0^*-a_1^*)}-\gamma}+1}$ and $y_i = \frac{1}{1-\frac{\gamma e^{\frac{1}{2}\zeta(a_0^*-a_1^*)^2+\sqrt{\zeta}(a_1^*-a_0^*)}}{\gamma-1}} + \gamma - 1$. Evaluating $\eta(a_0^*, y_i)\phi(\eta(a_0^*, y_i))$ at the critical points yields values $-\frac{1}{\sqrt{2\pi e}}$ and $\frac{1}{\sqrt{2\pi e}}$. Further, since the integral is over an interval of length 1 with uniform density, the integral has a lower bound of $-\frac{1}{\sqrt{2\pi e}}$. Substituting this into the second-order condition yields the condition

$$\frac{\partial^2 u(a;\theta)}{\partial a_i^2} + \frac{\zeta\Psi}{2\alpha}\frac{1}{\sqrt{2\pi e}} \leq 0,$$

yielding the condition $\zeta \leq -\frac{2\alpha\sqrt{2\pi e}}{\Psi}\frac{\partial^2 u(a;\theta=0)}{\partial a_i^2}$.

Since the second-order condition is negative at the equilibrium effort choice, it is a maximum. Further, this is the only maximum by concavity of the utility function. Therefore, such an optimal policy must be unique. Indeed, this is the unique equilibrium because

pooling equilibria cannot exist. Pooling can be ruled out by noticing that, in any pooling equilibrium, the probability of reelection is not a function of the choice variable (i.e, it is a constant). The solution to the problem in that case is the leader's ideal point, contradicting pooling. □