

Global Public Goods Provision, Information Dissemination, and Domestic Politics*

Justin Melnick[†]

August 28, 2024

Abstract

Flexible international institutions recently rely on leaders making voluntary contributions to public goods with reports of these contributions disseminated publicly. I study a setting where leaders invest costly effort into providing global public goods and are accountable to a domestic audience and an international organization. I contrast a model of domestic accountability, in which leaders can exert effort to signal competence to voters, with a model where the organization provides effort recommendations based on reported willingness to contribute. Paradoxically, investments into public goods are smaller under the organization because the institution's publicization of reports provides voters with information about leader type, breaking the accountability chain and disincentivizing leaders' investments. The model highlights a novel mechanism in the study of international cooperation that emanates from microfounding cooperation within domestic politics, namely the tradeoff between leaders' desire to signal to domestic publics and their reticence to be screened by international organizations.

Word Count: 7362 (text) + 774 (references) + 1999 (appendix) = 10135 (total)

*I am grateful to Pat Egan, Sanford Gordon, Layna Mosley, Peter Rosendorff, Tara Slough, Alastair Smith, Aleksandra Conevska, Kun Heo, Carolina Torreblanca, the NYU IR Workshop, the NYU Dissertation Seminar, the Harvard Political Economy Workshop, the SWIPER summer reading group, EITM 2023, the Washington University in St. Louis Formal Theory Conference, and the LSE-NYU Political Economy Conference for insightful comments and feedback.

[†]Ph.D. Candidate, New York University. melnickj@nyu.edu.

To facilitate the provision of global public goods, international institutions have developed increasingly flexible institutional arrangements focused on voluntary national commitments and international transparency mechanisms. These institutional features, like pledge-and-review frameworks (Harstad 2023a;b), allow political leaders to determine the extent to which they wish to comply with international rules; the institution aggregates and disseminates information about each state’s willingness to cooperate. What does international cooperation look like under such an institution?

As an example, consider the Paris Agreement to reduce greenhouse gas emissions.¹ The Paris framework allows countries to propose their own emissions reduction targets, implemented via national investments into mitigation policy. The agreement contains no formal enforcement mechanism at the international level: the United Nations Framework Convention on Climate Change (UNFCCC) simply collects and disseminates information about nations’ willingness to mitigate through “nationally determined contributions.” By contrast, Paris’s predecessor, the Kyoto Protocol, failed because it imposed “top-down,” legally binding emissions reduction targets onto industrialized nations and levied fines for noncompliance, which lead several countries to delay ratification and others to withdraw from the agreement altogether (Victor 2011; Keohane and Oppenheimer 2016); Kyoto ultimately fell victim to common political instabilities that perpetuate the underprovision of global public goods (Keohane and Victor 2016). The introduction of voluntary commitments and public reporting of these commitments into the realm of climate governance sought to address Kyoto’s shortcomings and ameliorate the problem of the “climate dead end” (Stern 2007).

This paper studies a setting in which a political leader is accountable to a domestic audience and is also party to an international organization with features like the Paris Agreement. Leaders voluntarily reveal their willingness to invest in global public goods and subsequently exert costly effort into contributions. The organization makes such reports public. Conventional wisdom on the structure of international agreements would contend that these features should increase cooperation due to increased flexibility (Rosendorff and Milner 2001) and transparency (Dai 2002). Contrarily, I demonstrate that these features have deleterious effects on leaders’ investments into public goods because the agreement’s design inspires less ambitious commitments through its effects on domestic politics.

By microfounding leader incentives to contribute to public goods within a model of domestic electoral accountability, I highlight a novel tradeoff that captures a tension between international cooperation and domestic political survival. While international agreements that seek to elicit leaders’ private information

¹Other comparable international institutions 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 Human Rights Council, and the International Labor Organization’s Complaint Procedure (Raiser, Çalı and Flachsland 2022).

intend to *screen* leaders based on their willingness to contribute to global public goods, domestic political motivations may incentivize leaders to *signal* competence to audiences at home through policy implementation. When international organizations publicize reports about leaders' willingness to contribute to public goods, it breaks any signaling incentives that would have encouraged greater investments. More generally, this setting corresponds to one in which an agent exerts costly effort in order to signal type to a principal, who decides whether to retain or replace the agent. Retention incentives driven by adverse selection motivate the agent to exert effort. However, when faced with a transparency mechanism that attempts to screen the agent's type, the agent's incentives to signal type to the principal dissipate and therefore effort attenuates.

To study how an international agreement's informational effects alter the prospects for cooperation, I compare equilibria of two models. In the first, I study the interaction between a leader and a voter without any overarching international institution. Leaders vary in their willingness to contribute to global public goods, their private type, captured by their marginal costs of exerting effort into investments (competence). The voter observes an imperfect signal of effort, which is informative of the leader's competence. Public goods provision forms the basis of an accountability relationship between the leader and the voter (cf. [Ashworth, Bueno de Mesquita and Friedenber](#) 2017); the leader's reelection prospects are thus conditioned on observable policy outcomes. Higher realizations of the voter's signal represent more "successful" policy outcomes, which on average reflect a leader who invested greater efforts into providing public goods. Leaders therefore find it politically advantageous to invest in public goods as it signals competence, here net of any collective action considerations that play into global public goods provision ([Ostrom 1990](#)).

The second model introduces an international organization seeking to maximize global investments into public goods. I consider a setting in which leaders report their type to the organization, representing their intentions of voluntary national commitments to the public good; the organization subsequently publicizes these reports. The core insight is that, because information about leader competence is also domestically relevant for voters, the agreement's design of aggregating and disseminating information detracts from leaders' incentives to exert effort rather than enhance them. Leaders face a tension in their incentives to exert effort because their private information is relevant to multiple audiences ([Farrell and Gibbons 1989](#)). Hence, with an agreement in place, the provision of public goods is expected to be lower than in a world without international cooperation.

The international organization detracts from leader effort because of its informational effects on domestic politics. The agreement's transparency mechanisms provide information about leader type and break the accountability relationship between leaders and voters by resolving uncertainty about the leader's competence.

Consequently, voter assessments of a leader's willingness to invest in public goods are no longer conditioned on leader effort, hence leaders' electoral incentives to exert effort dissipate.

Related Literature

Principally, this paper contributes to work elucidating international organizations' role of disseminating information to enhance cooperation (Keohane 1984; Dai 2005). When combined with theories of reputation-building, the provision of information by institutions may lead to cooperative outcomes in the long run (Milgrom, North and Weingast 1990). In this paper, I focus on the international provision of information to domestic audiences. Scholars have demonstrated that international institutions may motivate good governance by disseminating information about government performance (Kelley and Simmons 2015; 2019) with the hopes of disciplining government behavior (Besley and Burgess 2002; Hollyer, Rosendorff and Vreeland 2015). These studies contend that greater information provision will lead to better governance because information should strengthen the accountability channel between leaders and their publics; however, this paper joins other work that demonstrates how information from international institutions can worsen governance outcomes by distorting leaders' incentives within these domestic relationships (e.g., Hollyer and Rosendorff 2011; Bisbee et al. 2019).

This paper also examines the role of transparency in agency relationships. Transparency can worsen governance outcomes within accountability channels (Prat 2005; Stasavage 2007; Fox and Van Weelden 2012) because the direct provision of information decreases accountability (Besley 2006). This model reaffirms this finding. In the domestic context, the desire to remain in office motivates leaders to exert greater effort into providing public goods, as doing so is more likely to realize more successful policy outcomes, which reflect desirably on leader type and thus increase reelection prospects. However, transparency at the international level unravels this accountability relationship because the publicity of reports about competence nullify the electoral incentives to deepen cooperation.

Finally, this model connects to other formal treatments of agreements with voluntary commitments and public reporting, like pledge-and-review frameworks. I opt for a mechanism design framework to study cooperation, similar to the literature on international climate cooperation (Harrison and Lagunoff 2017; Slechten 2020; McAllister and Schnakenberg 2022). Additionally, Harstad (2023 *a*; *b*) studies a more general pledge-and-review bargaining setup and analyzes the optimal proposal when countries only propose their own share to a common project. I innovate on these models by microfounding leaders' incentives to cooperate within a model of domestic politics, and in so doing document new obstacles to cooperation.

The paper’s analysis proceeds by comparing two models. In the first, leaders invest in costly effort into providing global public goods in the shadow of domestic politics without any overarching international institution (the domestic politics game). This establishes the insight that leaders may be incentivized to invest effort for domestic electoral considerations. In the second, I introduce cooperation whereby an international organization makes effort recommendations for leaders to carry out and characterize the optimal recommendations given leaders’ political constraints (the international cooperation game). I demonstrate how cooperation unravels through the intersection of public reporting and domestic electoral incentives.

Domestic Politics Game

I begin by considering a model in which leaders exert effort into reducing emissions without any structured international cooperation.

Setup

Consider a strategic interaction between $n > 2$ countries that engage in policymaking on the provision of a global public good. In each country (indexed by i) there is a leader and a representative voter. Leader i has a private type, $\theta_i \in \{\underline{\theta}, \bar{\theta}\}$ with $0 < \underline{\theta} < \bar{\theta}$. Let the common prior be $P(\theta_i = \underline{\theta}) = q$. Only leader i knows her type; all voters and all other leaders only know the prior. Leaders’ types represent their competence, or the ease with which they can produce public goods. Think of competence as a marginal cost: leaders of type $\underline{\theta}$ have lower costs of producing public goods and therefore will be more willing to invest in them, hence they are competent. By contrast, leaders of type $\bar{\theta}$ face greater costs and are subsequently less willing to invest, deemed incompetent.²

Given their types, leaders choose a policy $a_i \in [0, \omega]$. These policies represent effort allocated toward providing the public good, where $\omega > 0$ is some maximum feasible effort level. Refer to the vector of effort levels exerted by each country as $a = (a_1, \dots, a_n)$. A higher value of a_i is consistent with more ambitious effort and thus a greater commitment to public goods provision on behalf of country i . Let the utility over policy given leader i ’s type be $u(a; \theta_i)$. I will assume that the function $u(a; \theta_i)$ is twice continuously differentiable in a_i , increasing and strictly concave. Further, let the cross-partial of effort a_i and type θ_i

²This cost captures three considerations. First, investing in global public goods may have political consequences for leaders because these policies could create domestic winners and losers. Potential opposition to these policies may be politically costly, and some leaders are more willing than others to incur these costs. Second, leaders vary in their *ex ante* ideological proclivity toward investing in international cooperative projects. Third, these investments may carry opportunity costs in the form of forsaking other policy agendas.

satisfy $\frac{\partial^2 u(a; \theta_i)}{\partial a_i \partial \theta_i} \leq 0$, consistent with the idea that θ_i is leader i 's marginal cost of producing the public good. Finally, I assume that $u(a; \theta_i)$ is additively separable between a_i and a_{-i} , the effort exerted by other countries besides i . This both eases exposition and allows me to concentrate on the effects of information that persist while holding fixed any collective action concerns.

Let the level of effort that satisfies $\frac{\partial u(a; \theta_i)}{\partial a_i} = 0$ be leader i 's "ideal point" level and denote it as \tilde{a}_i . Since $u(a; \theta_i)$ is continuous in a_i and is maximized over a compact interval, \tilde{a}_i exists. This is the level of effort that maximizes leader i 's utility over policymaking and serves as a convenient benchmark because it represents the level of effort that leaders would exert if the accountability mechanism described below was turned off.

Example 1 Consider the following utility function:

$$u(a; \theta_i) = A - \frac{\theta_i}{2} a_i^2, \quad (1)$$

where $A = \sum_i a_i$ denotes global effort levels. Leader i 's ideal effort is $\tilde{a}_i = \frac{1}{\theta_i}$. Competent leaders exert more effort than incompetent leaders, $\frac{1}{\underline{\theta}} > \frac{1}{\bar{\theta}}$.

The utility function in Example 1 captures several key considerations of international cooperation on public goods problems. Leaders benefit from aggregate effort investments: each leader is better off when nations allocate greater amounts of effort toward providing the public good. However, efforts are personally costly. This establishes the temptation for leaders to free ride off of the contributions of others, and as we shall see, to misrepresent one's type.

In addition to engaging with other nations, leaders are accountable to domestic publics when investing in global public goods. The voter sees a noisy signal of leader i 's effort,

$$K_i = a_i + \varepsilon_i,$$

where $\varepsilon_i \sim N(0, \frac{1}{\beta})$, and β is not too large.³

The signal literally implies that voters have imperfect information about leader effort, but could also represent the net value of policy, or the "success" in implementing policy, at the time of the election. Higher values of the signal are more likely to reflect more ambitious effort levels. Based on the realization of the signal, voter i determines whether to retain leader i or replace her, choosing $\rho_i \in \{0, 1\}$ where $\rho_i = 1$ denotes that the voter retains the leader and $\rho_i = 0$ denotes that he replaces her.

³This assumption is a sufficient condition to ensure that the incompetent leader's utility is concave at the extremum. See proof of Proposition 1 in the appendix.

Two elements comprise the voter's payoff. A voter receives benefits from having a competent leader in office, but, in addition to the selection problem, he 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 (a valence shock). Denote bias as $y_i \sim G(\cdot)$, where $G(\cdot)$ is a cumulative distribution function satisfying the monotone likelihood ratio property. For analytical simplicity, I work with $y_i \sim U[-\gamma, \gamma]$. The value of this bias is realized right before the voter makes his choice to retain the leader or not. To easily parameterize the voter's preferences for competence, suppose that the voter gets utility 1 from having a competent leader in power and 0 otherwise.⁴ If the voter replaces the incumbent leader, her replacement is of type θ_C such that $P(\theta_C = \underline{\theta}) = q$. The voter's utility function is thus

$$v(\rho_i; y_i) = \rho_i \left(\mathbb{1}_{\theta_i = \underline{\theta}} + y_i \right) + (1 - \rho_i) \mathbb{1}_{\theta_C = \underline{\theta}}.$$

The leader's payoff comprises both utility over public goods provision $u(a; \theta_i)$ and a benefit from remaining in power. If the voter retains the leader, the leader enjoys the value of holding office, $\Psi > 0$. The leader's payoff is thus

$$U(a, \rho_i; \theta_i) = u(a; \theta_i) + \rho_i \Psi.$$

The timing of the game is as follows:

1. Types θ_i are revealed to incumbent leaders.
2. Leaders simultaneously choose effort a_i .
3. Voters observe the signal K_i and bias y_i and retain or replace their leaders.
4. Payoffs realized. Game ends.

I examine Perfect Bayesian Equilibria. A leader's strategy maps type into an effort level, $a_i : \{\underline{\theta}, \bar{\theta}\} \rightarrow [0, \omega]$. The voter's strategy maps the realization of the signal and his bias into a retention rule, $\rho_i : \mathbb{R} \times [-\gamma, \gamma] \rightarrow \{0, 1\}$. The voter's retention rule is sequentially rational given conjectures about the leader's effort choice (since it is imperfectly observed) and beliefs about the leader's type. The leader's effort choices are sequentially rational with the voter's retention rule. Beliefs about type are determined by Bayes's rule.

⁴This payoff structure represents a reduced form version of a two-period model in which the voter has explicit preferences over the provision of public goods. Then, since the competent type can more easily provide public goods, there is a selection problem in favor of retaining competent types.

Analysis

By backward induction, first consider the voter's behavior. He adopts a straightforward retention rule in which voter i retains leader i if and only if his posterior belief about the leader's competence, as well as his bias, is greater than the probability that her replacement is competent,

$$\mu(K_i) + y_i \geq q,$$

where $\mu(K_i) = P(\underline{\theta}|K_i)$ is the voter's posterior belief about the leader's competence given the realized signal.

Since efforts are imperfectly observed, the voter in each country needs to have conjectures about the efforts chosen by each leader-type. Denote these by $(\hat{a}(\underline{\theta}), \hat{a}(\bar{\theta}))$. The voter retains the incumbent whenever the realization of policy is greater than or equal to some threshold, $K_i \geq \hat{K}$. Fixing the voter's bias y_i , the voter is thus exactly indifferent between retaining the incumbent leader and replacing her when

$$K(y_i) = \frac{\hat{a}(\underline{\theta}) + \hat{a}(\bar{\theta})}{2} + \frac{\log\left(\frac{(1-q)(q-y_i)}{q(1-q+y_i)}\right)}{\beta(\hat{a}(\underline{\theta}) - \hat{a}(\bar{\theta}))} \equiv \hat{K}(y_i).$$

This equality provides a relationship between the value of the signal that the voter needs to observe in order to retain the incumbent and a realized value of the voter's bias; the bias y_i affects the way he parses information about leader type. Increasing bias y_i decreases the relative value of selecting competent leaders and therefore creates a more permissive cutoff rule by which voters assess leader type. If $y_i > q$, then the voter's bias toward the incumbent dwarfs the value added of the signal, $K(y_i) \rightarrow -\infty$, and the voter would retain the incumbent for any value of K_i . Leaders who are already ingratiated with the voter thus survive in office regardless of how their effort into public goods provision translates into policy outcomes. Conversely, if $y_i < -1 + q$, then the voter's bias against the incumbent leader is large, $K(y_i) \rightarrow \infty$, and the leader would be replaced regardless of implemented policy.

Integrating out bias yields the probability that a leader of type θ_i is reelected,

$$\frac{1}{2\gamma} \int_{-1+q}^q \Phi(\sqrt{\beta}(a_i - \hat{K}(y_i))) dy + \frac{\gamma - q}{2\gamma}.$$

The first term represents the chance that the leader is reelected as a function of her investment into effort. The second term is not a function of this effort, because there is some probability that the leader is

reelected based on valence. A leader of type θ_i therefore has an expected utility of

$$EU_i(a; \theta_i) = u(a; \theta_i) + \left[\int_{-1+q}^q \Phi(\sqrt{\beta}(a_i - \hat{K}(y_i))) dy + \gamma - q \right] \frac{\Psi}{2\gamma}.$$

Leaders care about their utility derived from the policy benefits of public goods investments but also about the value of remaining in office, and choose a level of effort to maximize these priorities. Notice immediately that if electoral incentives are irrelevant ($\Psi = 0$), then leaders exert their ideal effort level \tilde{a}_i . When choosing optimal efforts, leaders must consider how their effort choices might reveal information about their type, and subsequently the effort's effect on electoral prospects. Leader-type θ_i 's effort choice satisfies the following first-order condition (applying the refinement that the voter's beliefs about effort choices are correct in equilibrium):

$$\underbrace{\frac{\partial u(a; \theta_i)}{\partial a_i} \Big|_{a_i = a^*(\theta_i)}}_{\text{marginal willingness to exert effort}} + \underbrace{\frac{\sqrt{\beta}\Psi}{2\gamma} \int_{-1+q}^q \phi\left(\sqrt{\beta}(a^*(\theta_i) - \hat{K}^*(y))\right) dy}_{\text{marginal benefit of office-holding}} = 0.$$

Leaders balance the marginal costs of exerting effort with their electoral returns and the marginal policy benefits from contribution. Manipulating this first-order condition yields several insights. First, leaders always behave more ambitiously compared to their ideal point, because doing so can signal competence to the voter. Effort is increasing in the value of electoral incentives Ψ , $\frac{\partial^2 u}{\partial a_i \partial \Psi} = \frac{\sqrt{\beta}}{2\gamma} \int_{-1+q}^q \phi\left(\sqrt{\beta}(a^*(\theta_i) - \hat{K}^*(y))\right) dy > 0$. It is clear that $a^*(\theta_i) > \tilde{a}(\theta_i)$ because the latter is the special case where $\Psi = 0$. Second, competent leaders still invest more than incompetent leaders. But now, the voter acts as a counter on the leader's temptations to pursue their less ambitious ideal points: exerting lower effort on average will signal to the voter that the leader is an incompetent type.

The results of the domestic politics game can be summarized in the following proposition (with proofs of all formal results in the appendix).

Proposition 1 *In the unique Perfect Bayesian equilibrium of the domestic politics game:*

- if $\Psi = 0$, leaders exert effort at their ideal points, $a^*(\theta_i) = \tilde{a}(\theta_i)$;
- competent leaders exert greater effort than incompetent leaders, $a^*(\underline{\theta}) > a^*(\bar{\theta})$;
- effort increases in the value of office-holding, $\frac{\partial}{\partial \Psi} a^*(\theta_i) \geq 0$;
- competent leaders are more likely to survive in office than incompetent leaders.

The domestic politics game provides microfoundations for how voters can use global public goods provision as an informational tool to evaluate relevant aspects of leader quality. Consequently, leaders exert effort into public goods provision keeping in mind how voters would update their beliefs from a signal that arises from these policy choices (cf. [Ashworth, Bueno de Mesquita and Friedenberg 2017](#)). Voters face a classic accountability problem in which they seek to sanction leaders for pursuing unambitious agendas by exerting control at the ballot box ([Ferejohn 1986](#)), subsequently rewarding leaders who choose policies that on average bring about better international cooperative outcomes ([Fearon 1999](#)).

Additionally, the model provides an explanation for why unilateral investments into public goods may be rational in the face of global collective action problems, as has been discussed about climate policy ([Aklin and Mildenberger 2020](#); [Kennard and Schnakenberg 2023](#)). If domestic publics can learn about leader quality from policy implementation, the opportunity to invest in public goods efforts and signal competence may be valuable for leaders.

International Cooperation Game

I now introduce an international organization (IO) into the model that recommends a level of effort for leaders to pursue. The IO might represent the UNFCCC which serves as an advisory body in collecting and disseminating information about nationally determined contributions within the framework of the Paris Agreement. The goal of the IO is to maximize commitments to providing the public good.

Setup

The IO interacts with leaders by soliciting information about each leader's type θ_i , which is done in the form of a report, and recommends efforts based on the reports. This can be thought of as the IO attempting to entice leaders to exert effort toward implementing (on average) better policy outcomes, conditional on the report a leader submits. These reports and subsequent recommendations are then made public to all other players. This organization has no enforcement power on its own, and serves purely to aggregate and disseminate private information, similar to Paris and other institutions with voluntary commitments and public dissemination of reports.

To characterize the optimal effort recommendations, I utilize the tools of mechanism design (cf. [Harrison and Lagunoff 2017](#); [McAllister and Schnakenberg 2022](#)). Think of the IO as a mechanism designer tailoring its optimal effort recommendations to leaders' reports. Let leader i 's report be $\hat{\theta}_i \in \{\underline{\theta}, \bar{\theta}\}$, with the IO's

corresponding recommendation being $x(\hat{\theta}_i)$. The IO designs $x(\hat{\theta}_i)$ to maximize the utilitarian objective function

$$V = \sum_i u(x(\hat{\theta}_i); \theta_i, \theta_{-i}).$$

If types were perfectly known, the IO would solve the problem above by prescribing the recommendation that solves the following first-order condition for each leader j :

$$\sum_i \frac{\partial u(x(\theta_j); \theta_j, \theta_{-j})}{\partial x_j} = 0,$$

which would be the “first best” effort level.

Example 1, continued *Given utility in Equation 1, the first best effort is $x(\theta_i) = \frac{n}{\theta_i}$.*

The first best exhibits normatively desirable properties. Recommendations are efficient, implementing a social optimum, and are “variable” in the sense that each leader invests effort according to their reported willingness. Should leaders report higher (lower) willingness to contribute to the public good, the IO recommends them a more (less) ambitious commitment (see [Harrison and Lagunoff \(2017\)](#) and [McAllister and Schnakenberg \(2022\)](#) for a discussion of variable recommendations in the context of carbon emissions abatement).

Since types are private information to leaders, they must decide whether or not to reveal this type truthfully to the IO. Applying the revelation principle, I examine the class of direct mechanisms that the IO could design such that its effort recommendations would be implementable in a Perfect Bayesian equilibrium of the domestic politics game analyzed above. The IO must therefore consider both the information and strategic constraints that leaders face. The information constraint requires that it is incentive compatible for leader i to reveal its true type θ_i , rather than lie and report type θ'_i , given that other leaders also submit truthful reports. The strategic constraint requires that it must be weakly profitable to obey the IO’s recommended level of effort.

The exercise at hand thus determines the IO’s optimal recommendations in the shadow of domestic electoral selection. Since leaders are symmetric, the IO only needs to design two recommendations: $x(\underline{\theta})$ and $x(\bar{\theta})$. The timing of the game is adjusted from the domestic politics game as follows:

1. Types θ_i are revealed to incumbent leaders.
2. Leaders simultaneously submit report of type $\hat{\theta}_i$ to the IO.

3. The IO recommends effort $x(\hat{\theta}_i)$ to leaders. Reports and recommendations are made public.
4. Voters observe the signal K_i and bias y_i and retain or replace their leaders.
5. Payoffs realized. Game ends.

A leader's strategy is now a report of its type, $\sigma_i : \{\underline{\theta}, \bar{\theta}\} \rightarrow \{\underline{\theta}, \bar{\theta}\}$, which the IO uses to determine an effort recommendation. Reports satisfy incentive compatibility and obedience constraints, to be defined formally below.

Comments on the Model

The mechanism design approach to modeling international cooperation imposes greater structure and thus warrants further discussion of additional modeling assumptions.

The leader's announcement. The leader's announcement of her willingness to contribute to public goods (competence) is akin to a cheap talk message. It is costless to send and need not be truthful. This message is analogous to the submission of nationally determined contributions in institutions like the Paris Agreement, or other transparency-enhancing procedures that elicit information about national capabilities to provide global public goods. As we shall see, incentive compatibility constraints provide conditions under which such information revelation would be truthful about leader type.

The IO's effort recommendations. Leaders make a report $\hat{\theta}_i$ of their type, which maps to a recommended effort level $x(\hat{\theta}_i)$. The IO's effort choice is analogous to $a(\theta_i)$ in the domestic politics game: the IO recommends the effort needed for the leader to implement policies to achieve the targets laid out in their voluntary commitments. The recommendation along with the subsequent obedience constraint ensures that this level of effort is individually rational for the leader.

To determine effort recommendations, the IO behaves like a utilitarian social planner, meaning that the IO hopes to realize the socially optimal effort investments given what leaders report about their abilities to contribute.⁵ However, leaders' domestic political constraints are crucial because they dictate truthful revelation of type and obedience of the IO's recommendations.

⁵The specific form of the IO's objective function is not needed to produce the main results.

The publicity of reports. I assume that when the IO receives self-reports from leaders, it disseminates this information worldwide. This assumption corresponds with the possibility for “naming and shaming” (e.g., [Hafner-Burton 2008](#); [Tingley and Tomz 2022](#)) by international and domestic audiences alike. The role of the IO in this model is to provide information to leaders and voters to clarify the uncertainty around θ_i . Consequently, if the mechanism is incentive compatible, voters have perfect information about leader type through the IO’s reporting.

The model considers an informational environment in which domestic voters can update their beliefs about leader type based on the IO’s dissemination of information. I study this setting because in institutions like the Paris Agreement, nationally determined contributions are announced and disseminated publicly. Studying variations in this informational environment may yield an institutional design that is capable of supporting greater levels of public goods investments in equilibrium, and is left for future research.

The obedience constraint. Since the bulk of the analysis considers the ramifications of information revelation, the leader’s incentive compatibility constraints are of central importance. The obedience constraint matters because institutions with voluntary commitments and public reporting often have no punishment mechanism since leaders propose their own level of compliance. It is often these types of mechanisms that discipline cooperation in other theories of international cooperation, typically modeled as repeated games (e.g., [Downs and Rocke 1995](#); [Rosendorff and Milner 2001](#)). Hence, the obedience constraint allows us to consider what levels of effort leaders would be willing to implement *ex post*, as the institution cannot compel them through any type of punishment mechanism.

Analysis

I begin with a benchmark case in which leaders face no electoral incentives ($\Psi = 0$), and demonstrate that recommendations are “compressed,” meaning they do not vary by reported type. I then reintroduce electoral incentives to demonstrate the main result: that the IO’s role in disseminating information about leader type discourages investment into public goods relative to a world without the IO by breaking the accountability chain between leaders and voters.

Without electoral benefits, the strategic tension that leaders face resembles the classic incentive to misrepresent marginal costs of effort. Declaring competence by pledging $\hat{\theta}_i = \underline{\theta}$ and exerting greater effort is a public good and all nations are better off if leader i contributes more, but it is costly to leader i to announce and implement such a policy. Leaders have incentives to lie about, and in particular, understate, their com-

petence. Leaders can either report their type θ_i truthfully, or they can mimic the other type of leader that could have been realized in their country, $\theta'_i \neq \theta_i$. Let $u(\hat{\theta}_i; \theta_i, \theta_{-i})$ be the utility that leader i receives when she is of type θ_i , reports $\hat{\theta}_i$ to the IO, and other nations are of type θ_{-i} . Incentive compatibility requires

$$u(\theta_i; \theta_i, \theta_{-i}) \geq u(\theta'_i; \theta_i, \theta_{-i}) \quad \forall \theta_i \in \{\underline{\theta}, \bar{\theta}\}, \quad \forall \theta'_i \in \{\bar{\theta}, \underline{\theta}\}.$$

To satisfy incentive compatibility, the utility that leaders receive from telling the truth about their type θ_i and receiving recommendation $x(\theta_i)$ must be weakly greater than the utility from reporting type $\theta'_i \neq \theta_i$ and receiving recommendation $x(\theta'_i)$. It is clear that incompetent leaders have no incentives to misreport and mimic competent leaders, as doing so would allocate them a tighter commitment to the public good than they would *ex ante* prefer. However, competent leaders may benefit from misreporting in order to avoid greater costs associated with more effort, thus mimicking incompetent leaders. Consequently, every leader would be tempted to report $\hat{\theta}_i = \bar{\theta}$, and the IO fails to screen competent leaders from incompetent leaders. To satisfy incentive compatibility constraints, the IO must prescribe a “compressed” effort recommendation to each leader regardless of their reported willingness to contribute. The IO’s recommendations are type-invariant within countries (cf. [Harrison and Lagunoff 2017](#); [McAllister and Schnakenberg 2022](#)).⁶ Counterintuitively, competent leaders obstruct efficient international cooperation.

Lemma 1 *Suppose $\Psi = 0$. Any incentive compatible mechanism is “compressed:” $x(\underline{\theta}) = x(\bar{\theta})$.*

Example 1, continued *The compressed recommendation takes the form*

$$x^*(\hat{\theta}_i) = \frac{n}{q\underline{\theta} + (1 - q)\bar{\theta}}.$$

What happens if leaders are both party to the agreement but also value domestic office? A key jump from the domestic politics game into the international cooperation game is that leader reports are public to voters. Moreover, if the agreement is incentive compatible such that it induces truthful revelation of leader type, voters learn the types of their leaders with certainty through these reports. The voter’s posterior belief μ about leader competence is either 1 or zero. Subsequently, whether leader i survives in office no longer depends on the signal K_i . The voter retains the leader if $y_i \geq q - \mu$ which occurs with probability $\frac{\gamma - q + \mu}{2\gamma}$.

⁶Because countries are symmetric (the values of $\underline{\theta}$, $\bar{\theta}$, and q are constant across countries), “within-country compression” and “across-country compression” are indistinguishable in the model but are conceptually distinct. Across-country compression means that every country receives the same effort recommendation. This behavior is reminiscent of the broader-deeper tradeoff ([Downs, Rocke and Barsoom 1998](#)) or the law of the least ambitious program ([Hovi, Ward and Grundig 2015](#)). If countries were not symmetric, across-country compression would not hold, but within-country compression would.

Information constraints for competent and incompetent leaders respectively now reflect this (suppressing dependence on the truthful reporting of all other nations θ_{-i}):

$$\begin{aligned}
\underbrace{u(\underline{\theta}; \underline{\theta})}_{\text{policy payoff for truthfully reporting } \theta_i = \underline{\theta}} + \underbrace{\frac{1}{2\gamma}\Psi}_{\text{net electoral benefit for reporting } \theta_i = \underline{\theta}} &\geq \underbrace{u(\bar{\theta}; \underline{\theta})}_{\text{policy payoff for reporting } \theta_i = \bar{\theta}} \\
\underbrace{u(\bar{\theta}; \bar{\theta})}_{\text{policy payoff for truthfully reporting } \theta_i = \bar{\theta}} &\geq \underbrace{u(\underline{\theta}; \bar{\theta})}_{\text{policy payoff for reporting } \theta_i = \underline{\theta}} + \underbrace{\frac{1}{2\gamma}\Psi}_{\text{net electoral benefit for reporting } \theta_i = \underline{\theta}}.
\end{aligned}$$

When leaders care about remaining in office, they must account for the fact that their submitted reports have downstream domestic consequences. If reports are incentive compatible, then the IO resolves the uncertainty regarding each leader's type. Any leader who reports to be competent would be reelected with probability $\frac{\gamma-q+1}{2\gamma}$; any leader who reports to be incompetent would be reelected with probability $\frac{\gamma-q}{2\gamma}$.

It is immediately clear that compressed recommendations are never incentive compatible for incompetent leaders, because they could always do better by claiming to be competent and thus increase their chances of reelection. Therefore, it must be the case that $x^*(\underline{\theta}) > x^*(\bar{\theta})$, or that competent leaders must always invest more effort than incompetent leaders, which follows from the monotonicity of leader type in policy utility. Moreover, since incompetent leaders win reelection with a lower probability, they are compensated through a more lenient effort recommendation. Conversely, because competent types benefit electorally from truthful revelation, they also shoulder greater costs by exerting more effort into the public good.

Introducing electoral incentives complicates intuition about the conditions under which leaders would report truthfully. Competent leaders may mimic incompetent leaders to incur fewer contribution costs, but incompetent leaders may mimic competent leaders to increase their reelection chances. It is thus not clear *ex ante* as to which type's incentive constraint would bind. Instead, a range on Ψ can be found such that both constraints would be simultaneously satisfied. This range is

$$2\gamma\left(u(\bar{\theta}; \underline{\theta}) - u(\underline{\theta}; \underline{\theta})\right) \leq \Psi \leq 2\gamma\left(u(\bar{\theta}; \bar{\theta}) - u(\bar{\theta}; \underline{\theta})\right).$$

This range defines the extent to which leaders value electoral benefits relative to utility from exerting effort into public goods. The lower bound on Ψ comes from the competent leader's constraint. By mimicking an incompetent type, a competent leader incurs fewer costs from exerting effort, but misreporting comes at the expense of decreased electoral odds. If electoral incentives are large enough, then the competent leader

finds it sufficient to forego the utility from a more lenient effort recommendation and report her true type, which is exactly the opposite of what she would do if $\Psi = 0$. Conversely, incentive compatibility for the incompetent leader requires that the value of domestic office is not too large, establishing the upper bound. If electoral incentives are hefty, incompetent leaders would be incentivized to mimic competent leaders, even if it means exerting more effort than they would prefer to increase the chances of remaining in power.

Evidently, what can be supported as an incentive compatible effort level depends on how leaders trade off the value of providing international public goods and maintaining domestic political office. To the extent that it can, the IO seeks to design efforts $x(\underline{\theta})$ and $x(\bar{\theta})$ to maximize the range of electoral incentives such that truthful revelation is incentive compatible. However, since leaders cannot exert infinite amounts of effort into providing public goods, the IO's tools in optimizing this tradeoff are limited. This is particularly difficult when thinking about cases in which leaders care primarily about domestic political survival, $\Psi \rightarrow \infty$. If survival concerns become too important, then no incentive compatible mechanism exists.

Lemma 2 *There exists a such $\bar{\Psi}$ such that no incentive compatible mechanism exists if $\Psi > \bar{\Psi}$.*

Example 1, continued *The first best could be sustained as an incentive compatible recommendation if $\frac{n\gamma(\bar{\theta}-\underline{\theta})(n\underline{\theta}+\bar{\theta}(n-2))}{\bar{\theta}^2\underline{\theta}} \leq \Psi \leq \frac{n\gamma(\bar{\theta}-\underline{\theta})(n\bar{\theta}+\underline{\theta}(n-2))}{\bar{\theta}\underline{\theta}^2}$. No incentive compatible mechanism exists if $\Psi > \frac{\gamma(\bar{\theta}\omega-1)^2}{\bar{\theta}}$.*

Lemma 2 establishes the limit on the tradeoff between investing effort into public goods and holding office. Incompetent leaders find it incentive compatible to report truthfully only if Ψ is not too high, because reporting competence is electorally valuable. If Ψ is larger than this upper bound, it is never politically valuable for incompetent leaders to tell the truth. Incompetent leaders would rather pretend to be competent in order to increase their electoral odds, violating incentive compatibility.⁷

Thus far, I have only examined whether leaders find it incentive compatible to reveal their true type to the IO. I now winnow the set of incentive compatible recommendations to those that satisfy *ex post* obedience constraints: conditional on the public reports of type, leaders must be willing to implement their recommended policies rather than some other profitable deviation. Such a constraint is appropriate as it embodies the idea that no leader ever needs to be a part of the international agreement. The obedience

⁷If $\Psi > \bar{\Psi}$, leaders always report to the IO that they are competent, $\hat{\theta}_i = \underline{\theta}$, regardless of their true type (similar to babbling (cf. Crawford and Sobel 1982)). The IO recommends the largest mitigation investment that could be supported by some type of pooling, as in Lemma 1. However, this recommendation is not implementable because it fails the obedience constraint. See Corollary 1 in the appendix.

constraint thus requires

$$\underbrace{u(x^*(\theta_i); \theta_i)}_{\text{policy payoff for obeying IO as type } \theta_i} + \underbrace{\frac{\gamma - q + \mu}{2\gamma} \Psi}_{\text{electoral payoff for obeying IO as type } \theta_i} \geq \max_d \underbrace{u(d; \theta_i, \theta_{-i})}_{\text{policy payoff for maximal deviation as type } \theta_i} + \underbrace{\frac{\gamma - q + \mu}{2\gamma} \Psi}_{\text{electoral payoff for deviating as type } \theta_i}.$$

The publicity of leader reports nullifies the electoral benefits from exerting costly effort into public goods provision for all leaders *ex post*, regardless of their type. Since the IO publicizes leaders' reports, resolving uncertainty around type, leaders' electoral odds are no longer connected to the effort that they exert into international cooperation. This removes any incentives for leaders to comply with particularly ambitious investments into public goods because the IO washes out the leader's accountability relationship with the voter. The only effort level that satisfies leader i 's obedience constraint is her ideal point $\tilde{a}(\theta_i)$. The revelation of information deactivates the electoral mechanism through which leaders exert effort in the hopes of increasing their electoral odds.

Lemma 3 *Recommendations $x^*(\theta_i)$ satisfy leaders' obedience constraints if and only if $x^*(\theta_i) = \tilde{a}(\theta_i)$.*

The result follows directly from the fact that the IO reveals electorally relevant information to voters by making information related to leader type publicly available. The IO breaks the accountability chain between leaders and voters, so leaders have no incentives to make more ambitious commitments to international public goods than their ideal efforts. Paradoxically, leaders would always pursue greater investments into public goods without the framework of an international agreement.

Proposition 2 *In an equilibrium implemented by the international cooperation game:*

- *leaders exert effort at their ideal points, $x^*(\theta_i) = \tilde{a}(\theta_i)$, only if $\Psi \in [0, \bar{\Psi}]$;*
- *competent leaders exert greater effort than incompetent leaders, $x^*(\underline{\theta}) > x^*(\bar{\theta})$;*
- *the value of office-holding has no effect on effort, $\frac{\partial}{\partial \Psi} x^*(\theta_i) = 0$;*
- *competent leaders are more likely to survive in office than incompetent leaders;*
- *expected global effort is less than in the domestic politics game, $E[X^*] \leq E[A^*]$.*

In equilibrium, the IO can not entice leaders to invest any effort greater than their ideal points because it publicizes information about leader type and washes out the electoral incentives to exert effort into providing public goods. This recommendation is always incentive compatible for competent leaders, but the ideal

effort level is only incentive compatible for incompetent leaders if $\Psi \leq \bar{\Psi}$. The features of the international institution counteract the domestic incentives for more ambitious action in providing public goods, leading to a suboptimal global provision of effort.

Since the beginning of the pledge-and-review process, experts have assessed that nations have failed to “deliver on their promises” to meet the Paris goal of limiting warming to below 1.5-2°C, calling for more ambitious climate action (UNEP 2021; UNFCCC 2023). Equilibrium behavior in the model would suggest that expected observed policy outcomes fall short of countries’ achievement of Paris goals because the effort they put into meeting those goals is minimal. Indeed, more ambitious mitigation efforts may not have been credible by design of the international institutional structure.

Figure 1 summarizes the key strategic implications across both the domestic politics game and the international cooperation game, where various effort choices are plotted as a function of electoral incentives Ψ . The solid black dots characterize leaders’ ideal efforts. Introducing electoral incentives, as in the domestic politics game, increases effort because leaders now have the ability to signal competence to a domestic audience. This is illustrated by the solid purple lines in which optimal effort $a^*(\theta_i)$ in the domestic politics game is increasing in the value of Ψ .

In the international cooperation game, the IO strives to maximize international commitments to public goods investments. The normative ideal would be for each leader to exert effort at the social optimum, which is captured by the purple crosses in Figure 1. Contrasted with their ideal points, leaders internalize their externalities in the first best. However, the IO relies on leaders to self-report their willingness to contribute, which constrains the ability for the IO to implement this social optimum. Indeed, since leaders have private information about their willingness to contribute, the IO must incentivize truthful revelation of this information.

If there are no electoral incentives, introducing informational constraints compresses the IO’s recommended effort levels to the black triangle. But when leaders also care about reelection, the compressed recommendation is not possible. Moreover, leaders face strategic constraints on top of their informational constraints. Each leader must be willing to exert the recommended effort, and the compressed recommendation fails this obedience criterion. The equilibrium of the international cooperation game is therefore illustrated by the dashed grey lines, in which leaders simply implement their ideal points. The IO fails to motivate leaders to exert more effort, and worsens investments relative to what they would be in the domestic politics game in which leaders have electoral incentives to behave more ambitiously.

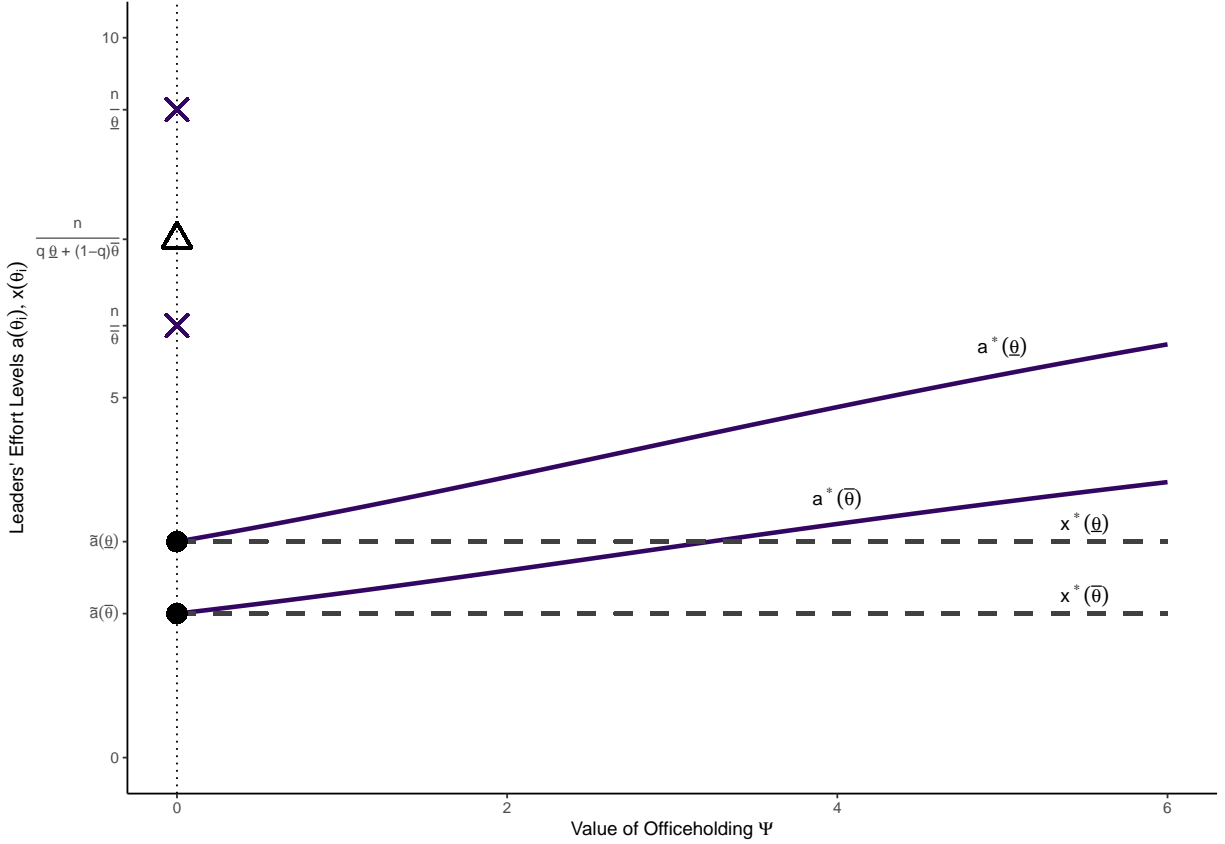


Figure 1: Optimal Efforts in the Domestic Politics and International Cooperation Games

Figure constructs equilibrium outcomes using Example 1. The solid black points are leaders' ideal efforts, $\tilde{a}(\theta_i) = \frac{1}{\theta_i}$. The solid purple lines are leaders' equilibrium efforts in the domestic politics game as a function of the value of holding office, $a^*(\theta_i)$. The purple crosses are leaders first best efforts in the international cooperation game, $\frac{n}{\theta_i}$. The black triangle is the IO's compressed recommendation in the international cooperation game, $\frac{n}{q\theta + (1-q)\bar{\theta}}$. The dashed grey lines are the IO's optimal recommendations in the domestic politics game, $x^*(\theta_i) = \frac{1}{\theta_i}$.

Who Joins?

Why would leaders be compelled to join an institution with voluntary commitments and public reporting? Clearly, if some leaders can enhance their electoral odds through the signaling mechanism, then being a part of such an institution detracts from this goal. Moreover, on policy grounds, the present analysis would suggest that leaders would be better off without international cooperation to facilitate the provision of global public goods. I augment the analysis to consider a membership stage in which leaders can decide whether or not they want to join the IO, and provide conditions on the parameter space in which a pooling equilibrium where both types of leaders join can be sustained.

Proposition 3 *Competent leaders always join the agreement. There exists a threshold $\bar{\gamma}$ such that incompetent leaders join the agreement if and only if $\gamma > \bar{\gamma}$.*

The incentives to join the agreement depend on how leaders benefit electorally from information revelation. Intuitively, competent leaders have the most to gain from joining the agreement: they do not need to over-invest in terms of the effort needed to signal type, and they are electorally rewarded. However, incompetent types may wish to play the domestic politics game to enhance their electoral odds. Although it means exerting more effort into providing public goods than they would ideally prefer, incompetent leaders are less likely to survive in office when party to the agreement. This tradeoff boils down the salience of the policy outcome generated by global public goods provision relative to other issues that the voter cares about when at the ballot box, parameterized by γ . The IO resolves the voter's selection problem, which detracts from incompetent leaders' electoral odds relative to what would happen in the domestic politics game. If γ is large, incompetent leaders could still win the election based on her popularity on other electorally relevant issues, i.e., if the voter's bias y_i toward the incumbent is high.

This result can help to rationalize the broad membership that institutions like the Paris Agreement enjoy. While increasing in salience over time (Egan, Konisky and Mullin 2022), climate change's effects on electoral outcomes continue to be fairly minor. If γ is large, even those unwilling to pursue bold climate reforms may find the stakes of joining the agreement low. Indeed, such leaders benefit because they can exert less effort into mitigation relative to what they would do in an equilibrium without the agreement ($x^*(\theta_i) < a^*(\theta_i)$), and they can salvage their electoral odds through their popularity on other issues. Moreover, large γ also implies a weakening of the accountability mechanism between leaders and their domestic publics on the issue of global public goods provision, as policy outcomes through this channel are not as politically dispositive. This may clarify why leaders in places with weaker accountability relationships, for example in less democratic

societies, are willing to join institutions like the Paris Agreement.

Conclusion

This paper zoomed into the effects of two features found in contemporary international cooperative institutions: voluntary commitments and public reporting of information about those commitments. I show that, due to a multi-layered agency problem, leaders have fewer incentives to exert effort when party to such an international institution. This is because leaders want to simultaneously signal competence to their voters, thereby incentivizing costly investments into global public goods, but also face incentives to understate their willingness to contribute at the international level. Leaders face a tension in their incentives to exert effort because their private information is relevant to multiple audiences: signaling to voters increases effort, while screening by the organization decreases effort. Paradoxically, the transparency facilitated by the agreement provides voters with more precise information about their leaders, which demotivates leaders in using effort as a signal of leader type.

References

- Aklin, Michaël and Matto Mildenerger. 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.
- Ashworth, Scott and Ethan Bueno de Mesquita. 2006. “Monotone Comparative Statics for Models of Politics.” *American Journal of Political Science* 50(1):214–231.
- Ashworth, Scott, Ethan Bueno de Mesquita and Amanda Friedenberg. 2017. “Accountability and Information in Elections.” *American Economic Journal: Microeconomics* 9(2):95–138.
- Besley, Timothy. 2006. *Principled Agents?: The Political Economy of Good Government*. Oxford: Oxford University Press.
- Besley, Timothy and Robin Burgess. 2002. “The Political Economy of Government Responsiveness: Theory and Evidence from India.” *The Quarterly Journal of Economics* 117(4):1415–1451.
- 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.
- Crawford, Vincent P. and Joel Sobel. 1982. “Strategic Information Transmission.” *Econometrica* 50(6):1431–1451.
- 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.
- Downs, George and David M. Rocke. 1995. *Optimal Imperfection?: Domestic Uncertainty and Institutions in International Relations*. Princeton: Princeton University Press.
- Downs, George W., David M. Rocke and Peter N. Barsoom. 1998. “Managing the Evolution of Multilateralism.” *International Organization* 52(2):397–419.
- Egan, Patrick J, David M Konisky and Megan Mullin. 2022. “Ascendant Public Opinion: The Rising Influence of Climate Change on Americans’ Attitudes about the Environment.” *Public Opinion Quarterly* 86(1):134–148.

- Farrell, Joseph and Robert Gibbons. 1989. "Cheap Talk with Two Audiences." *The American Economic Review* 79(5):1214–1223.
- 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.
- Ferejohn, John. 1986. "Incumbent Performance and Electoral Control." *Public Choice* 50(1/3):5–25.
- Fox, Justin and Richard Van Weelden. 2012. "Costly transparency." *Journal of Public Economics* 96(1):142–150.
- Hafner-Burton, Emilie M. 2008. "Sticks and Stones: Naming and Shaming the Human Rights Enforcement Problem." *International Organization* 62(4):689–716.
- Harrison, Rodrigo and Roger Lagunoff. 2017. "Dynamic Mechanism Design for a Global Commons." *International Economic Review* 58(3):751–782.
- Harstad, Bård. 2023a. "Pledge and review bargaining." *Journal of Economic Theory* 207(C).
- Harstad, Bård. 2023b. "Pledge-and-Review Bargaining: from Kyoto to Paris." *The Economic Journal* 133(651):1181–1216.
- Hollyer, James R. and B. Peter Rosendorff. 2011. "Why Do Authoritarian Regimes Sign the Convention Against Torture? Signaling, Domestic Politics and Non-Compliance." *Quarterly Journal of Political Science* 6(3–4):275–327.
- Hollyer, James R., B. Peter Rosendorff and James Raymond Vreeland. 2015. "Transparency, Protest, and Autocratic Instability." *American Political Science Review* 109(4):764–784.
- Hovi, Jon, Hugh Ward and Frank Grundig. 2015. "Hope or Despair? Formal Models of Climate Cooperation." *Environmental & Resource Economics* 62(4):665–688.
- 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 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. 2016. "Cooperation and discord in global climate policy." *Nature Climate Change* 6:570–575.
- Keohane, Robert O. and Michael Oppenheimer. 2016. "Paris: Beyond the Climate Dead End through Pledge and Review?" *Politics and Governance* 4(3):142–151.
- McAllister, Jordan H. and Keith E. Schnakenberg. 2022. "Designing the Optimal International Climate Agreement with Variability in Commitments." *International Organization* 76(2):469–486.
- Milgrom, Paul and Chris Shannon. 1994. "Monotone Comparative Statics." *Econometrica* 62(1):157–180.
- Milgrom, Paul R., Douglass C. North and Barry R. Weingast. 1990. "The Role of Institutions in the Revival of Trade: The Law Merchant, Private Judges, and the Champagne Fairs." *Economics & Politics* 2(1):1–23.
- Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Dallas, TX: Cambridge University Press.
- Prat, Andrea. 2005. "The Wrong Kind of Transparency." *The American Economic Review* 95(3):862–877.
- 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.
- Rosendorff, B. Peter and Helen V. Milner. 2001. "The Optimal Design of International Trade Institutions: Uncertainty and Escape." *International Organization* 55(4):829–857.
- Slechten, Aurélie. 2020. "Environmental Agreements under Asymmetric Information." *Journal of the Association of Environmental and Resource Economists* 7(3):455–481.
- Stasavage, David. 2007. "Polarization and Publicity: Rethinking the Benefits of Deliberative Democracy." *Journal of Politics* 69(1):59–72.
- Stern, Nicholas. 2007. *The Economics of Climate Change: The Stern Review*. 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.

UNEP. 2021. Emissions Gap Report 2021: The Heat is On - A World of Climate Promises Not Yet Delivered. Technical report UNEP Copenhagen Climate Centre Nairobi: .

UNFCCC. 2023. Technical dialogue of the first global stocktake. Technical report United Arab Emirates: .

Victor, David G. 2011. *Global Warming Gridlock: Creating More Effective Strategies for Protecting the Planet*. Cambridge: Cambridge University Press.

Appendix

All formal results from the main text are reproduced and proven here.

Proposition 1 *In the unique Perfect Bayesian equilibrium of the domestic politics game:*

- if $\Psi = 0$, leaders exert effort at their ideal points, $a^*(\theta_i) = \tilde{a}(\theta_i)$;
- competent leaders exert greater effort than incompetent leaders, $a^*(\underline{\theta}) > a^*(\bar{\theta})$;
- effort increases in the value of office-holding, $\frac{\partial}{\partial \Psi} a^*(\theta_i) \geq 0$;
- competent leaders are more likely to survive in office than incompetent leaders.

We will prove Proposition 1 with a series of claims.

Claim 1 *The unique equilibrium of the domestic politics game is characterized by a double (\underline{a}, \bar{a}) , which represents leader-type θ_i 's policy choices that forms a Bayesian Nash equilibrium given the policy choices of leader-types in other countries θ_{-i} . The voter in country i retains the leader if and only if $K_i \geq \hat{K}(y_i)$.*

Proof of Claim 1: Begin by noting that because countries are symmetric, all leaders with type $\underline{\theta}$ will choose the same policy, as will all leaders with type $\bar{\theta}$.

Voter i adopts a decision rule in which he retains the leader if and only if

$$P(\underline{\theta}|K_i) + y_i \geq q.$$

The voters in each country need to have conjectures about the policies chosen by each leader-type. Denote these by $(\hat{a}(\underline{\theta}), \hat{a}(\bar{\theta}))$. Posterior beliefs about leader i 's type given the observed value of the signal are

$$P(\underline{\theta}|K_i) = \frac{q\phi(\sqrt{\beta}(K_i - \hat{a}(\underline{\theta})))}{q\phi(\sqrt{\beta}(K_i - \hat{a}(\underline{\theta}))) + (1 - q)\phi(\sqrt{\beta}(K_i - \hat{a}(\bar{\theta})))}.$$

Conditional on some value of his bias y_i , the voter is thus exactly indifferent between retaining the incumbent leader and replacing her when

$$K(y_i) = \frac{\hat{a}(\underline{\theta}) + \hat{a}(\bar{\theta})}{2} + \frac{\log\left(\frac{(1-q)(q-y_i)}{q(1-q+y_i)}\right)}{\beta(\hat{a}(\underline{\theta}) - \hat{a}(\bar{\theta}))}.$$

The likelihood ratio $\frac{\phi(\sqrt{\beta}(K_i - \hat{a}(\underline{\theta})))}{\phi(\sqrt{\beta}(K_i - \hat{a}(\bar{\theta})))}$ is increasing in the signal K_i . Therefore, the voter in country i retains his leader if and only if $K_i \geq \hat{K}$. Also note that if $y_i > q$ then $K(y_i) \rightarrow -\infty$ and if $y_i < -1 + q$ then $K(y_i) \rightarrow \infty$. The threshold \hat{K} that the voter uses to reelect the incumbent is

$$\hat{K}(y_i) = \begin{cases} \infty & y_i < -1 + q \\ \frac{\hat{a}(\underline{\theta}) + \hat{a}(\bar{\theta})}{2} + \frac{\log\left(\frac{(1-q)(q-y_i)}{q(1-q+y_i)}\right)}{\beta(\hat{a}(\underline{\theta}) - \hat{a}(\bar{\theta}))} & -1 + q < y_i < q \\ -\infty & y_i > q. \end{cases}$$

Clearly, this means that if $y_i > q$ the leader is retained with probability 1 and if $y_i < -1 + q$ 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 $-1 + q < y_i < q$. Therefore, the probability of reelection can be decomposed into two terms. If $y_i > q$, the leader survives with probability 1, which occurs with $P(y_i > q) = \frac{\gamma - q}{2\gamma}$. Second, if $-1 + q < y_i < q$, the leader survives with probability $\Phi(\sqrt{\beta}(a_i - \hat{K}(y_i)))$. Therefore, the total probability of survival in office is

$$\frac{1}{2\gamma} \int_{-1+q}^q \Phi(\sqrt{\beta}(a_i - \hat{K}(y_i))) dy + \frac{\gamma - q}{2\gamma}.$$

Leader i maximizes the following expected utility:

$$EU_i(a; \theta_i) = u(a; \theta_i) + \left[\int_{-1+q}^q \Phi(\sqrt{\beta}(a_i - \hat{K}(y_i))) dy + \gamma - q \right] \frac{\Psi}{2\gamma}.$$

For type θ_i , the first-order condition is

$$\frac{\partial u(a; \theta_i)}{\partial a_i} + \frac{\sqrt{\beta}\Psi}{2\gamma} \int_{-1+q}^q \phi\left(\sqrt{\beta}\left(a_i - \frac{\hat{a}(\underline{\theta}) + \hat{a}(\bar{\theta})}{2} - \frac{\log\left(\frac{(1-q)(q-y_i)}{q(1-q+y_i)}\right)}{\beta(\hat{a}(\underline{\theta}) - \hat{a}(\bar{\theta}))}\right)\right) dy = 0.$$

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

$$\frac{\partial u(a; \theta_i)}{\partial a_i} + \frac{\sqrt{\beta}\Psi}{2\gamma} \int_{-1+q}^q \phi\left(\sqrt{\beta}\left(\frac{a^*(\underline{\theta}) + a^*(\bar{\theta})}{2} - \frac{\log\left(\frac{(1-q)(q-y_i)}{q(1-q+y_i)}\right)}{\beta(a^*(\underline{\theta}) - a^*(\bar{\theta}))}\right)\right) dy = 0.$$

Because leaders/countries are symmetric, there are 2 equations in 2 unknowns. Solving these equations yield optimal effort levels $(a^*(\underline{\theta}), a^*(\bar{\theta}))$. To confirm that the equilibrium policy choices are a maximum, I take the second-order condition. Define $\eta(a_i, y_i) = \sqrt{\beta}\left(a_i - \frac{\hat{a}(\underline{\theta}) + \hat{a}(\bar{\theta})}{2} - \frac{\log\left(\frac{(1-q)(q-y_i)}{q(1-q+y_i)}\right)}{\beta(\hat{a}(\underline{\theta}) - \hat{a}(\bar{\theta}))}\right)$. Using the fact that

$\frac{d}{da}\phi(\eta) = -\eta\phi(\eta)\frac{\partial\eta}{\partial a}$, the second-order condition is

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

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

Now consider the second-order condition for type $\bar{\theta}$. Note that $\eta(a^*(\bar{\theta}), y_i) = \frac{a^*(\bar{\theta}) - a^*(\underline{\theta})}{2} - \frac{\log\left(\frac{(1-q)(q-y_i)}{q(1-q+y_i)}\right)}{\beta(\hat{a}(\underline{\theta}) - \hat{a}(\bar{\theta}))}$ need not be positive. A sufficient condition to show that the equilibrium effort $a^*(\bar{\theta})$ is a maximum is to find a lower bound on the integral. Differentiating $\eta(a^*(\bar{\theta}), y_i) \phi(\eta(a^*(\bar{\theta}), y_i))$ with respect to y_i yields the critical points $y_i = \frac{q-1}{\frac{1}{qe^{\frac{1}{2}b(a^*(\bar{\theta}) - a^*(\underline{\theta}))^2 + \sqrt{b}(a^*(\bar{\theta}) - a^*(\underline{\theta}))} + 1}}$ and $y_i = \frac{1}{1 - qe^{\frac{1}{2}b(a^*(\bar{\theta}) - a^*(\underline{\theta}))^2 + \sqrt{b}(a^*(\bar{\theta}) - a^*(\underline{\theta}))}} + q - 1$. Evaluating $\eta(a^*(\bar{\theta}), y_i) \phi(\eta(a^*(\bar{\theta}), y_i))$ 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

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

yielding the condition $\beta \leq -\frac{2\gamma\sqrt{2\pi e}}{\Psi} \frac{\partial^2 u(a; \bar{\theta})}{\partial a_i^2}$.

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, $\tilde{a}(\theta_i)$ in which $\tilde{a}(\underline{\theta}) \neq \tilde{a}(\bar{\theta})$, contradicting pooling. ■

Claim 2 *Equilibrium efforts are decreasing in θ , increasing in Ψ , and decreasing in γ .*

Proof of Claim 2: By the tools of monotone comparative statics (Milgrom and Shannon 1994; Ashworth and Bueno de Mesquita 2006), I conclude that $\frac{\partial a_i^*}{\partial \theta_i} \leq 0$, $\frac{\partial a_i^*}{\partial \gamma} \leq 0$, and $\frac{\partial a_i^*}{\partial \Psi} \geq 0$ for any a_i^* that maximizes leader i 's expected utility. For any θ and any Ψ at the equilibrium choice of a_i ,

$$\frac{\partial^2 EU_i}{\partial a_i \partial \theta_i} = \frac{\partial^2 u(a; \theta_i)}{\partial a_i \partial \theta_i} \leq 0.$$

$$\frac{\partial^2 EU_i}{\partial a_i \partial \Psi} = \frac{\sqrt{\beta}}{2\gamma} \int_{-1+q}^q \phi\left(\sqrt{\beta}(a_i - \hat{K}(y_i))\right) dy > 0.$$

$$\frac{\partial^2 EU_i}{\partial a_i \partial \gamma} = -\frac{\sqrt{\beta}\Psi}{2\gamma^2} \int_{-1+q}^q \phi\left(\sqrt{\beta}(a_i - \hat{K}(y_i))\right) dy < 0.$$

■

Proof of Proposition 1: Existence and uniqueness of the equilibrium is established in Claim 1. That $a^*(\underline{\theta}) > a^*(\bar{\theta})$ follows from Claim 2 because policy choices are increasing in θ . That $\frac{\partial a_i^*}{\partial \Psi} \geq 0$ is immediate from Claim 2. Since the probability of surviving is increasing in K_i and K_i is increasing in effort a_i , competent leaders are likely to survive in office than incompetent leaders because $a^*(\underline{\theta}) > a^*(\bar{\theta})$. ■

Lemma 1 Suppose $\Psi = 0$. Any incentive compatible mechanism is “compressed:” $x(\underline{\theta}) = x(\bar{\theta})$.

Proof of Lemma 1: Since types are private information, the IO’s objective function is

$$V = \max_{\{x(\underline{\theta}), x(\bar{\theta})\}} \sum_i q \left[u(x(\underline{\theta}); \theta_i, \theta_{-i}) \right] + (1-q) \left[u(x(\bar{\theta}); \theta_i, \theta_{-i}) \right].$$

Since countries are symmetric and utility is additively separable in θ_{-i} for each leader i , we can rewrite the problem as

$$V = \max_{\{x(\underline{\theta}), x(\bar{\theta})\}} n \left[qu(\underline{\theta}; \underline{\theta}) + (1-q)u(\bar{\theta}; \bar{\theta}) \right],$$

The IO wishes to maximize V subject to the incentive constraints

$$u(\theta_i; \theta_i, \theta_{-i}) \geq u(\theta'_i; \theta_i, \theta_{-i}) \quad \forall \theta_i \in \{\underline{\theta}, \bar{\theta}\}, \quad \forall \theta'_i \in \{\bar{\theta}, \underline{\theta}\}.$$

The monotonicity of leader utility in θ_i requires that $x(\underline{\theta}) \geq x(\bar{\theta})$. Notice that the incompetent type would never mimic the competent type, as doing so would lead her to receive a more stringent recommendation than she would prefer. The competent type, however, could choose to mimic the incompetent type, receiving a less ambitious recommendation. Therefore, the incentive constraint of the competent type must bind, or

$$u(\underline{\theta}; \underline{\theta}) = u(\bar{\theta}; \underline{\theta}).$$

Again by symmetry, the incentive constraint for all competent types of all countries bind simultaneously.

Using this and the competent type's incentive constraint further simplifies the IO's objective function to

$$V = \max_{\{x(\underline{\theta}), x(\bar{\theta})\}} n \left[qu(\bar{\theta}; \underline{\theta}) + (1 - q)u(\bar{\theta}; \bar{\theta}) \right],$$

where all leaders report $\theta_i = \bar{\theta} \forall \theta_i$. The solution to this problem is compressed. This means that the IO assigns the same policy regardless of reported type, $x^*(\underline{\theta}) = x^*(\bar{\theta})$. Such a policy is incentive compatible because it yields the same utility regardless of whether leader i reports $\hat{\theta}_i = \underline{\theta}$ or $\hat{\theta}_i = \bar{\theta}$. To see this, notice that if not, $x^*(\underline{\theta}) \neq x^*(\bar{\theta})$, monotonicity requires $x(\underline{\theta}) > x(\bar{\theta})$ for $\underline{\theta} < \bar{\theta}$. Finally, because of the concavity of the leader's utility function, we have that for $\underline{\theta} < \bar{\theta}$, the competent type's interim expected utility is greater if it mimics the incompetent type, which contradicts incentive compatibility. Thus any solution is compressed. ■

Lemma 2 *There exists a such $\bar{\Psi}$ such that no incentive compatible mechanism exists if $\Psi > \bar{\Psi}$.*

Proof of Lemma 2: Immediate from the incentive constraint of the incompetent type. The incompetent type reports truthfully if and only if

$$u(\bar{\theta}; \bar{\theta}, \theta_{-i}) \geq u(\underline{\theta}; \bar{\theta}, \theta_{-i}) + \frac{1}{2\gamma}\Psi \Leftrightarrow \Psi \leq 2\gamma \left(u(\bar{\theta}; \bar{\theta}, \theta_{-i}) - u(\underline{\theta}; \bar{\theta}, \theta_{-i}) \right).$$

The LHS is increasing in Ψ and the RHS is constant in Ψ so the constraint is satisfied whenever $\Psi \leq \bar{\Psi}$. ■

Lemma 3 *Recommendations $x^*(\theta_i)$ satisfy leaders' obedience constraints if and only if $x^*(\theta_i) = \tilde{a}(\theta_i)$.*

Proof of Lemma 3: The obedience constraint of leader i with type θ_i is

$$u(x^*(\theta_i); \theta_i) + \frac{\gamma - q + \mu}{2\gamma}\Psi \geq \max_d u(d; \theta_i, \theta_{-i}) + \frac{\gamma - q + \mu}{2\gamma}\Psi.$$

The solution to the right-hand side is that the optimal deviation is $d = \tilde{a}(\theta_i)$, which requires that $x^*(\theta_i) = \tilde{a}(\theta_i)$ and the constraint is met with equality. ■

Proposition 2 *In an equilibrium implemented by the international cooperation game:*

- *leaders exert effort at their ideal points, $x^*(\theta_i) = \tilde{a}(\theta_i)$, only if $\Psi \in [0, \bar{\Psi}]$;*
- *competent leaders exert greater effort than incompetent leaders, $x^*(\underline{\theta}) > x^*(\bar{\theta})$;*

- the value of office-holding has no effect on effort, $\frac{\partial}{\partial \Psi} x^*(\theta_i) = 0$;
- competent leaders are more likely to survive in office than incompetent leaders;
- expected global effort is less than in the domestic politics game, $E[X^*] \leq E[A^*]$.

Proof of Proposition 2: That the implementable policy is the ideal effort follows from Lemma 3, since it is the only policy that would satisfy obedience constraints. It is also incentive compatible for a competent leader to invest her ideal effort, because in equilibrium her constraint requires $\Psi \geq 2\gamma(u(\tilde{a}(\bar{\theta}); \underline{\theta}) - u(\tilde{a}(\underline{\theta}); \underline{\theta}))$. Since $u(\tilde{a}(\underline{\theta}); \underline{\theta})$ maximizes policy utility, the constraint is negative and therefore always satisfied. From Lemma 2, the incompetent type's effort is incentive compatible if $\Psi \leq \bar{\Psi}$. It is immediate that $\tilde{a}(\underline{\theta}) > \tilde{a}(\bar{\theta})$ and that $\frac{\partial \tilde{a}(\theta_i)}{\partial \Psi} = 0$ from the definition of the leader's utility over effort. Aggregate expected effort in the international cooperation game is $E[X^*] = n(q\tilde{a}(\underline{\theta}) + (1-q)\tilde{a}(\bar{\theta}))$, and by Claim 2, the equilibrium policies of the domestic politics game are greater than the ideal policies. ■

Proposition 3 *Competent leaders always join the agreement. There exists a threshold $\bar{\gamma}$ such that incompetent leaders join the agreement if and only if $\gamma > \bar{\gamma}$.*

Proof of Proposition 3: Define $p(\theta_i) = \int_{-1+q}^q \Phi(\sqrt{\beta}(a^*(\theta_i) - \hat{K}^*(y_i))) dy$. The competent leader prefers to join the agreement if

$$u(\tilde{a}(\underline{\theta}); \theta_i) + \frac{\gamma - q + 1}{2\gamma} \Psi \geq u(a^*(\underline{\theta}); \theta_i) + \frac{\gamma - q + p(\underline{\theta})}{2\gamma} \Psi,$$

which always holds because $u(\tilde{a}(\underline{\theta}); \theta_i) > u(a^*(\underline{\theta}); \theta_i)$ and $p(\underline{\theta}) < 1$.

The incompetent leader prefers to join the agreement if

$$u(\tilde{a}(\bar{\theta}); \theta_i) + \frac{\gamma - q}{2\gamma} \Psi \geq u(a^*(\bar{\theta}); \theta_i) + \frac{\gamma - q + p(\bar{\theta})}{2\gamma} \Psi \Leftrightarrow u(\tilde{a}(\bar{\theta}); \theta_i) \geq u(a^*(\bar{\theta}); \theta_i) + \frac{p(\bar{\theta})}{2\gamma} \Psi.$$

Observe that the LHS is constant in γ and, by the envelope theorem, the RHS is decreasing in γ . Therefore there is $\bar{\gamma}$ where the incompetent leader is indifferent between joining and not joining. She joins the agreement if and only if $\gamma > \bar{\gamma}$. ■

Corollary 1 *Suppose $\Psi > \bar{\Psi}$. The compressed effort recommendation fails the obedience constraint.*

Proof of Corollary 1: Since $\Psi > \bar{\Psi}$, both types of leader i report $\theta_i = \underline{\theta}$, getting policy utility $u(\underline{\theta}; \theta_i)$. Recall a leader of type θ_i pursuing effort a_i gets reelected with probability

$$\frac{1}{2\gamma} \left[\gamma - q + \int_{-1+q}^q \Phi \left(\sqrt{\beta} (a_i - \frac{\hat{a}(\underline{\theta}) + \hat{a}(\bar{\theta})}{2}) - \frac{\log \left(\frac{(1-q)(q-y_i)}{q(1-q+y_i)} \right)}{\beta(\hat{a}(\underline{\theta}) - \hat{a}(\bar{\theta}))} \right) dy \right].$$

There are two cases. Suppose that the voter believes that, as hypothesized, leaders are pooling across types. This implies that $\hat{a}(\underline{\theta}) = \hat{a}(\bar{\theta})$, meaning the final term of the reelection probability is $\Phi(-\infty) = 0$. The obedience constraint therefore requires

$$u(\underline{\theta}; \theta_i) + \frac{\gamma - q}{2\gamma} \Psi \geq \max_d u(d; \theta_i, \theta_{-i}) + \frac{\gamma - q}{2\gamma} \Psi$$

As in Lemma 3, the unique solution is for leaders to deviate to $\tilde{a}(\theta_i)$, the utility-maximizing effort, contradicting obedience of the compressed recommendation.

In the second case, suppose the voter believes that leaders are separating, $\hat{a}(\underline{\theta}) \neq \hat{a}(\bar{\theta})$. Denote the compressed recommendation as \hat{x} . Then the obedience constraint is

$$u(\underline{\theta}; \theta_i) + \frac{\Psi}{2\gamma} \left[\gamma - q + \int_{-1+q}^q \Phi \left(\sqrt{\beta} (\hat{x} - \frac{\hat{a}(\underline{\theta}) + \hat{a}(\bar{\theta})}{2}) - \frac{\log \left(\frac{(1-q)(q-y_i)}{q(1-q+y_i)} \right)}{\beta(\hat{a}(\underline{\theta}) - \hat{a}(\bar{\theta}))} \right) dy \right] \geq$$

$$\max_d u(d; \theta_i, \theta_{-i}) + \frac{\Psi}{2\gamma} \left[\gamma - q + \int_{-1+q}^q \Phi \left(\sqrt{\beta} (d - \frac{\hat{a}(\underline{\theta}) + \hat{a}(\bar{\theta})}{2}) - \frac{\log \left(\frac{(1-q)(q-y_i)}{q(1-q+y_i)} \right)}{\beta(\hat{a}(\underline{\theta}) - \hat{a}(\bar{\theta}))} \right) dy \right].$$

The right-hand side of the constraint is simply leader i 's utility in the equilibrium of the domestic politics game, $d = a^*(\theta_i)$. Then, by Proposition 1, we know that the unique equilibrium of the domestic politics game requires that leaders separate, $a^*(\underline{\theta}) \neq a^*(\bar{\theta})$. Hence pooling on the compressed recommendation cannot be optimal. ■