Online Appendix for: Global Public Goods Provision, Information Dissemination, and Domestic Politics

Contents

A Comments on the Model

The mechanism design approach to modeling international cooperation (Harrison and Lagunoff 2017; McAllister and Schnakenberg 2022) 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. 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 game without the agreement: 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—although this specific functional form is not necessary to produce the main result—meaning that the IO hopes to realize the socially optimal effort investments given what leaders report about their abilities to contribute. However, leaders' domestic political constraints are crucial because they dictate truthful revelation of type and obedience of the IO's recommendations.

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.

B Formal Proofs

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

Proposition 1 *In the unique PBE without the international agreement:*

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

I prove Proposition 1 with a series of claims.

Claim 1 *The unique equilibrium of the game without the agreement is characterized by a double* $(\underline{a}, \overline{a})$*, 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 $\overline{\theta}$.

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 $(\hat{a}(\underline{\theta}), \hat{a}(\overline{\theta}))$. Posterior beliefs about leader i's type given the observed value of the signal are

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

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

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

The likelihood ratio $\frac{\phi((K_i - \hat{a}(\underline{\theta})))}{\phi((K_i - \hat{a}(\overline{\theta}))}$ 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 $K(y_i) \to -\infty$ and if $y_i < -1 + q$ then $K(y_i) \to \infty$. The threshold $\hat{K}$ that the voter uses to reelect the incumbent is

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

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 $\Phi((a_i - \hat{K}(y_i)))$. Therefore, the total probability of survival in office is

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

Leader i maximizes the following expected utility:

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

For type θ_i, the first-order condition is

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

Equilibrium requires that voters' conjectures are correct, $\hat{a}(\theta_i) = a^*(\theta_i)$, so this simplifies to

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

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

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

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

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

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

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

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 $\tilde{a}(\underline{\theta}) \neq \tilde{a}(\overline{\theta})$, contradicting pooling. $\blacksquare$

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,

$$\frac{\partial^2 EU_i}{\partial a_i \partial \theta_i} = \frac{\partial^2 u(a;\theta_i)}{\partial a_i \partial \theta_i} \leq 0.$$
$$\frac{\partial^2 EU_i}{\partial a_i \partial \Psi} = \frac{}{2\gamma} \int_{-1+q}^{q} \phi\Big((a_i - \hat{K}(y_i))\Big)\, dy > 0.$$
$$\frac{\partial^2 EU_i}{\partial a_i \partial \gamma} = -\frac{\Psi}{2\gamma^2} \int_{-1+q}^{q} \phi\Big((a_i - \hat{K}(y_i))\Big)\, dy < 0.$$

■

Proof of Proposition 1: Existence and uniqueness of the equilibrium is established in Claim 1. That $a^*(\underline{\theta}) > a^*(\overline{\theta})$ 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 $a^*(\underline{\theta}) > a^*(\overline{\theta})$. ■

Lemma 1 *There exists a $\bar{\Psi}$ such that no incentive compatible mechanism exists if $\Psi > \bar{\Psi}$.*

Proof of Lemma 1: Immediate from the incentive constraint of the incompetent type. The incompetent type reports truthfully if and only if

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

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

Lemma 2 *Recommendations $x^*(\theta_i)$ satisfy obedience constraints if and only if $x^*(\theta_i) = \tilde{a}(\theta_i)$.*

Proof of Lemma 2: 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 agreement:*

- *leaders truthfully report their types if $\Psi \in [0, \bar{\Psi}]$ and obey recommendations of their ideal points $x^*(\theta_i) = \tilde{a}(\theta_i)$;*
- *competent leaders exert greater effort than incompetent leaders, $x^*(\underline{\theta}) > x^*(\overline{\theta})$;*
- *the value of office-holding has no effect on effort, $\frac{\partial}{\partial \Psi} x^*(\theta_i) = 0$;*
- *expected global effort is less than in the game without the agreement.*

Proof of Proposition 2: That the implementable policy is the ideal effort follows from Lemma 2, 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 $\Psi \geq 2\gamma(u(\tilde{a}(\overline{\theta});\underline{\theta}) - u(\tilde{a}(\underline{\theta});\underline{\theta}))$. Since $u(\tilde{a}(\underline{\theta});\underline{\theta})$ maximizes policy utility, the constraint is negative and therefore always satisfied. From Lemma 1, the incompetent type's effort is incentive compatible if $\Psi \leq \bar{\Psi}$. It is immediate that $\tilde{a}(\underline{\theta}) > \tilde{a}(\overline{\theta})$ 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 $E[X^*] = n(q\tilde{a}(\underline{\theta}) + (1-q)\tilde{a}(\overline{\theta}))$, and by Claim 2, the equilibrium policies of the game without the agreement are greater than the ideal policies. ■

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

Proof of Proposition 3: Define $p(\theta_i) = \int_{-1+q}^{q} \Phi((a^*(\theta_i) - \hat{K}^*(y_i)))\ dy$. 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

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

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 $\gamma > \bar{\gamma}$. ■

Lemma 3 *Suppose $\Psi = 0$. Any incentive compatible mechanism is "compressed:" $x(\underline{\theta}) = x(\overline{\theta})$.*

Proof of Lemma 3: Since types are private information, the IO's objective function is

$$V = \max_{\{x(\underline{\theta}),\ x(\overline{\theta})\}} \sum_i q\Big[u(x(\underline{\theta}); \theta_i)\Big] + (1-q)\Big[u(x(\overline{\theta}); \theta_i)\Big].$$

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

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

The IO wishes to maximize V subject to the incentive constraints

$$u(x(\theta_i); \theta_i) \geq u(x(\theta_i'); \theta_i)\ \forall \theta_i \in \{\underline{\theta}, \overline{\theta}\},\ \forall \theta_i' \in \{\overline{\theta}, \underline{\theta}\}.$$

The monotonicity of leader utility in θ_i requires that $x(\underline{\theta}) \geq x(\overline{\theta})$. 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

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

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

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

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

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 $u(x(\underline{\theta});\theta_i)$. Recall a leader of type θ_i pursuing effort a_i gets reelected with probability

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

There are two cases. Suppose that the voter believes that, as hypothesized, leaders are pooling across types. This implies that $\hat{a}(\underline{\theta}) = \hat{a}(\overline{\theta})$, meaning the final term of the reelection probability is

$\Phi(-\infty) = 0$. The obedience constraint therefore requires

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

As in Lemma 2, 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, $\hat{a}(\underline{\theta}) \neq \hat{a}(\overline{\theta})$. Denote the compressed recommendation as $\hat{x}$. Then the obedience constraint is

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

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

The right-hand side of the constraint is simply leader i's utility in the equilibrium of the game without the agreement, $d = a^*(\theta_i)$. Then, by Proposition 1, we know that the unique equilibrium of the game without the agreement requires that leaders separate, $a^*(\underline{\theta}) \neq a^*(\overline{\theta})$. Hence pooling on the compressed recommendation cannot be optimal. ■

C Who Joins?

This section conditions when leaders would join the agreement described in the main text but is omitted due to space constraints. 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 $\overline{\gamma}$ such that incompetent leaders join the agreement if and only if $\gamma > \overline{\gamma}$.*

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 game without the agreement 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 without the agreement. 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.

References

Ashworth, Scott and Ethan Bueno de Mesquita. 2006. "Monotone Comparative Statics for Models of Politics." *American Journal of Political Science* 50(1):214–231.

Downs, George and David M. Rocke. 1995. *Optimal Imperfection?: Domestic Uncertainty and Institutions in International Relations*. Princeton: Princeton University Press.

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.

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.

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.

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

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.