

Strategic Obfuscation in International Benchmarking

Justin Melnick*

February 25, 2026

Abstract

International organizations increasingly rely on voluntary reporting to construct global performance indicators, such as international benchmarks, to motivate domestic governance reforms, with the expectation that information dissemination enhances domestic accountability. However, this paper demonstrates that such mechanisms can have unintended consequences for compliance with international goals. Using a formal model in which leaders simultaneously report to an international benchmarking organization and navigate domestic electoral incentives, the analysis reveals that information dissemination does not beget more ambitious governance reforms. Instead, leaders anticipate how publicized information will affect voter beliefs and, in equilibrium, systematically obfuscate their reported commitments. As a result, expected governance reforms are no greater in a world with international benchmarking than in a world without it. This study highlights a fundamental tradeoff between international good governance goals and domestic political survival, offering new insights into institutional design and the structural limits of self-reporting frameworks.

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

Over the last two decades, international organizations (IOs) have increasingly pivoted away from top-down enforcement in favor of voluntary reporting, peer review, and benchmarking. Across domains ranging from human rights to anti-corruption, frameworks such as the Open Government Partnership and the UN Convention Against Corruption require states to publicly disclose their intended policy goals, which the IO then aggregates and disseminates in the form of scorecards, rankings, or other metrics. The conventional wisdom underpinning the rise of these global performance indicators (GPIs) is that international transparency empowers domestic democratic accountability ([Kelley and Simmons 2015; 2019](#)). By providing voters and civil society with credible, standardized information about a government's commitments, IOs reduce uncertainty and incentivize leaders to deliver on public goods ([Dai 2002; Fang 2008](#)).

Empirical observations of these transparency regimes note that leaders frequently make ambitious international commitments to good governance, submitting robust national action plans or publicly embracing international benchmarks, only to subsequently fail to exert the costly political effort required to implement them. A growing critical literature notes that international benchmarking often leads to strategic manipulation, where states either substitute effort away from unmeasured areas ([Bisbee et al. 2019](#)) or focus on bureaucratic compliance rather than true reform ([Honig 2018; 2019](#)). While these studies importantly demonstrate that the existence of a benchmark distorts the allocation of state effort, they do not study the effects of GPIs on electoral accountability. Consequently, what remains under-theorized is how the introduction of a voluntary reporting platform, through which IOs craft GPIs, fundamentally alters the informational environment that domestic voters use to evaluate their leaders.

This note addresses a foundational question underlying this literature: can the provision of additional information through international transparency frameworks motivate electoral accountability? I argue that the answer, counterintuitively, is no. Previous scholarship about

the role that IOs play in information dissemination under-appreciates the strategic incentives of national leaders within self-reporting frameworks, who navigate a dual audience: the international community and their domestic electorate (Mitchell 1998; Haftendorn, Keohane and Wallender 1999; Dai 2005). The introduction of a voluntary reporting platform alters the domestic signaling environment that motivates investments into good governance. Because the information disseminated by these IOs relies on costless self-reporting, it is subject to strategic manipulation. Rather than enhancing the informational environment, the provision of this additional information induces strategic obfuscation, rendering the IO's signal uninformative and thus endogenously irrelevant for domestic electoral accountability.

To demonstrate this, I develop a formal model of domestic politics in the shadow of an international reporting mechanism. The model distinguishes between “aligned” leaders who share the median voter’s preference for governance reform and “misaligned” leaders who are beholden to special interests or engage in corruption and graft. Neither the IO nor the public can distinguish between these types *ex ante*, but the intent of the IO is to “screen” for leaders making ambitious commitments to good governance. These commitments form the raw data for GPIs and other international scorecards, which the IO publicizes. Domestic voters then use the standardized information disseminated by these GPIs to update their beliefs about their leader’s willingness to enact policies consistent with the public’s interest. Paradoxically, while GPIs could motivate greater domestic efforts by clarifying a state’s willingness to cooperate (Keohane 1984; Mitchell 2006; Awad and Riquelme 2024), they inadvertently deter such efforts when leaders fear this benchmarking will be used by domestic audiences to evaluate their alignment with the public’s interest.

The mechanism driving this result is that leaders construct their strategic reports based entirely on how voters will interpret the resulting IO scorecard. Because an ambitious GPI ranking signals that a leader shares the public’s preference for reform, misaligned leaders have strong electoral incentives to overcommit to good governance on paper while under-

delivering in practice.¹ Aligned leaders, who genuinely value reform and want to separate themselves from misaligned types, might attempt to signal their true type through even stronger reports. However, a chase-and-evade logic emerges (cf. Roberson 2006): misaligned and aligned leaders pool on the same ambitious reports, obscuring the voter's inferences. As a result, the GPI's public rankings become fundamentally uninformative of a leader's true preferences for reform.

The central finding is counterintuitive: leaders obfuscate in their disclosure of governance reforms precisely because the IO publicizes this information. Rather than demonstrate openness to global good governance goals through an ambitious pledge *ex ante*, leaders prefer to shift the burden of signaling to downstream policymaking where costly effort, though more expensive, offers a politically safer signal to domestic audiences. In equilibrium, the presence of the IO's benchmarking apparatus does not inspire any greater effort toward governance reform compared to a world without it. Interestingly, the problem does not depend on institutional strength but lies within institutional structure: IOs designed explicitly to disseminate information, and without proper enforcement power, can be strategically subverted by leaders seeking to maintain their domestic political office.

This paper's insight sits at the intersection of two large literatures in international relations: the value of flexibility in institutional design and the role of IOs as information disseminators. International agreements are frequently critiqued for their rigidity, as strict enforcement regimes are often poorly suited to accommodate diverse domestic circumstances (Rosendorff and Milner 2001; Rosendorff 2005; Johns 2014). In response, many contemporary institutions employ voluntary reporting and peer-review structures. These agreements are celebrated because they are maximally flexible (Linos and Pegram 2016), albeit often quite shallow (Downs, Rocke and Barsoom 1998; Gilligan 2004; Edry 2020). Typically,

¹This is similar to the finding in the international negotiations literature that incumbents have incentives to negotiate “weak treaties” to manage domestic audiences; see Buisseret and Bernhardt (2018); Battaglini and Harstad (2020); Melnick and Smith (2023).

they lack formal enforcement mechanisms entirely, relying instead on the aggregation and dissemination of a state's intended contributions to good governance.

Within this flexible architecture, I highlight a novel effect of domestic politics on international cooperation. Existing theories of information and cooperation largely treat domestic politics as a retrospective enforcement mechanism: IOs provide neutral information to voters about whether a leader has defected from an agreement, and voters punish those defectors (Mansfield, Milner and Rosendorff 2002; Dai 2005; Fang 2008; Chaudoin 2014). However, this approach generally occurs without modeling the leader's strategic disclosure of the very information that IOs disseminate. Domestic politics are not merely a constraint on compliance, but a central driver of what an IO knows, based on the voluntary reports it receives, and can subsequently broadcast to the public (e.g., Bayer and Crippa 2025). It is precisely the anticipated effect of this information dissemination on domestic political survival that stifles a leader's incentive to commit to bold reforms in the first place.

Consequently, this argument refines recent research examining the domestic political effects of international benchmarking and transparency. International institutions routinely issue global performance indicators designed to motivate better governance by exposing underperformance or highlighting best practices (Kelley and Simmons 2015; 2019). While one might expect that greater information provision strengthens the accountability channel between leaders and their publics, the publicization of information by IOs can have unintended consequences. Just as leaders might strategically substitute effort away from other priorities toward meeting a published goal (Bisbee et al. 2019), they may also strategically manipulate the reporting process itself. Extensive studies on transparency demonstrate that information disclosed by leaders, even when filtered through IOs, fundamentally alters voter beliefs and subsequent political support (Hollyer, Rosendorff and Vreeland 2015; 2018; 2019*a;b*). This paper builds on that foundation, arguing that within voluntary peer-review frameworks, leaders strategically co-opt the dissemination of information to preserve their domestic survival,

thereby neutralizing the institution’s ability to improve governance outcomes.

More broadly, this argument builds on a rich tradition in formal political economy demonstrating that increased transparency does not unambiguously improve democratic accountability. A robust literature highlights how certain forms of transparency or third-party reporting can create perverse incentives (e.g., [Prat 2005](#); [Ashworth and Shotts 2010](#)). When leaders know their actions, messages, or pledges are being directly monitored, they frequently alter their behavior to strategically manipulate the selection problem underlying electoral accountability. I extend this logic to the international arena, demonstrating how the specific informational architecture of self-reporting frameworks triggers these exact electoral distortions. Because the IO publicizes a leader’s intended commitments rather than just downstream outcomes, the institution inadvertently creates a platform for strategic obfuscation rather than enhance accountability.

Model

I develop a model of domestic politics in the shadow of an international reporting mechanism. There is an incumbent leader (she) and a representative, or median, voter (he).² The leader determines how much effort $a \geq 0$ to exert toward costly governance reforms (e.g., anti-corruption enforcement or bureaucratic modernization). The voter observes the informational environment and decides whether to retain the incumbent or replace her with a challenger, $r \in \{0, 1\}$.

Leaders vary in the extent to which they are influenced by corrupt or anti-reform special interests, which I parameterize as a marginal cost of pursuing reforms. “Aligned” leaders ($\theta = 1$) produce reform effort at marginal cost $\lambda_1 > 0$, while “misaligned” leaders ($\theta = 0$) act on behalf of special interests seeking to maintain the status quo, effectively raising their

²While framed around elections to ease exposition, the model applies equally to nondemocratic settings where a pivotal elite or electorate determines the leader’s survival in office ([Bueno de Mesquita et al. 2003](#)).

marginal costs of policy reform to λ_0 such that $\lambda_0 > \lambda_1$. Thus, aligned leaders are so named because they can exert effort more cheaply into good governance reforms that are favored by the public. Whether a leader is aligned or misaligned is her private information; the voter believes the incumbent is aligned with prior probability $\gamma \in (0, 1)$.

The incumbent leader's utility comprises a benefit from office-holding $\Psi > 0$ less the costs of pursuing effort toward reform. Without loss of generality, leaders obtain utility normalized to zero when out of office. The leader's utility while in office is

$$u_L(a; \theta) = r\Psi - \frac{\lambda_\theta}{2}a^2.$$

Observe that, in the absence of reelection incentives, leaders would not exert effort into good governance reforms, as such effort is costly.

The leader's reform effort generates a policy outcome, or quality of governance, which is observed by the voter, $x = a + \varepsilon$, where $\varepsilon \sim N(0, \frac{1}{\zeta})$ and $\zeta < \frac{\lambda_1 \sqrt{2\pi e}}{\Psi}$. We can think of ζ as a measure of “transparency,” which modulates the precision with which the voter perceives the mapping from the leader's effort to policy outputs. Substantively, this may represent “bureaucratic quality” ([Slough 2024](#)) or even the monitoring done by civil society groups about how a leader's effort into reform maps to the quality of governance. Higher values of the policy outcome are more likely to reflect more ambitious effort levels. The voter casts a ballot either for the incumbent leader or a potential challenger based on the realization of x . If the incumbent loses the election, the challenger assumes office and is aligned with probability γ .

The voter prefers to have aligned leaders in office, and so his payoff is

$$u_V(r) = r\theta + (1 - r)\theta_C,$$

where θ_C is the alignment of the challenger.³

To analyze the effect of international benchmarking, I consider an international organization that acts as an automated recipient and disseminator of intended governance commitments. Prior to exerting effort, the leader can submit a costless report $m \in \{0, 1\}$ to the IO. A report of $m = 1$ represents an ambitious commitment to governance reform, while $m = 0$ represents a weak commitment. The IO takes this self-reported data and publicizes it as a global performance indicator or scorecard. The resulting IO report is defined as $s = m + \nu$, where $\nu \sim N(0, \frac{1}{\tau})$. The parameter τ captures the institution's measurement capacity; a larger τ means the IO's scorecard is a more precise reflection of the leader's submitted commitment.⁴

In this setup, the IO is modeled as a non-strategic actor that mechanically disseminates the leader's report. While international organizations in some domains likely possess their own bureaucratic preferences ([Fang and Stone 2012](#)), their mandate within benchmarking frameworks is explicitly defined: to maximize global governance effort by publicizing state commitments. Because the IO's goal of fostering greater reform is relatively straightforward and relies on preserving the credibility of the benchmark, it lacks the incentive to strategically manipulate individual scorecards. This rationale lends credibility to the current theoretical representation. More importantly, abstracting away from IO agency provides the hardest test for the model's theoretical mechanism: as we shall see, it demonstrates that even when an IO faithfully and objectively executes its transparency mandate, the strategic survival incentives of domestic leaders are sufficient to neutralize the institution's informational value.

The timing of the game is as follows:

1. Nature draws the incumbent leader's type $\theta \in \{0, 1\}$, observed only by the leader.

³This is equivalent to a two-period model in which the voter's utility is parameterized over governance outcomes x .

⁴From a technical perspective, modeling the IO's report s as a continuous signal ensures non-degenerate beliefs about the leader's type on the equilibrium path, allowing for a clean characterization of the equilibrium without relying on complex refinements for off-path beliefs.

2. The leader sends a report $m \in \{0, 1\}$ to the IO. The IO disseminates the report $s = m + \nu$ to the public.
3. The leader exerts effort into reforms $a \in \mathbb{R}_+$.
4. The voter observes the IO report s and the domestic policy outcome $x = a + \varepsilon$, and decides whether to retain or replace the incumbent.
5. Payoffs realized. Game ends.

The solution concept is perfect Bayesian equilibrium (henceforth equilibrium). To isolate the effect of the international reporting mechanism on domestic politics, I contrast two institutional environments. First, I establish a baseline model in the absence of the IO, where voters must infer the leader's type solely from policy outcomes. Second, I introduce the IO reporting stage to analyze how the opportunity for public, costless signaling at the international level alters the leader's effort provision and the voter's electoral strategy.

Analysis

I begin by establishing the baseline equilibrium of the domestic interaction without the IO, and then contrast it with the augmented model where leaders submit intended governance plans to the international reporting mechanism. All formal proofs are relegated to the appendix.

Domestic Politics

As a benchmark, I solve the game when the leader acts independently and does not submit reports to an IO. First consider the voter's retention decision; he uses the policy outcome x to screen for aligned leaders. Let $\mu(x) = P(\theta = 1|x)$ be the voter's posterior belief that

the incumbent is aligned, and \hat{a}_θ be the voter's conjecture about the leader's effort. By Bayes's rule, the voter retains the incumbent if and only if his posterior belief is greater than the prior probability of drawing an aligned challenger, $\mu(x) \geq \gamma$, which reduces to a simple threshold rule: the voter reelects the incumbent if and only if

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

The leader then knows she will be reelected with probability

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

and selects a level of effort given by the first-order condition shown in Proposition 1.

Proposition 1. *In the unique equilibrium without the international organization, optimal efforts a_θ^* solve the first-order condition*

$$\underbrace{\sqrt{\zeta}\phi(\sqrt{\zeta}(\frac{a_1^* - a_0^*}{2}))\Psi}_{\text{marginal electoral benefit}} = \underbrace{\lambda_\theta a_\theta^*}_{\text{marginal policy costs}}.$$

The voter reelects the incumbent if and only if $x \geq \hat{x}(a^)$.*

To determine equilibrium effort levels, leaders balance the marginal costs of good governance reforms with the marginal electoral value of implementing reforms. As aligned leaders have lower marginal costs of effort, they naturally exert more effort than misaligned leaders ($a_1^* > a_0^*$). Thus, since greater effort translates into better governance outcomes on average, the voter is more likely to conclude that the leader is aligned upon observing a higher value of the policy outcome x . Consequently, better policy outcomes lead to greater odds of reelection.

This baseline establishes that the prospect of enhanced electoral odds motivates invest-

ments into good governance and provides a benchmark upon which we can compare any potential effects of international reporting.

International Reporting

Now consider the augmented model where the IO receives and disseminates the leader's self-reported reform commitments via a GPI or some other metric of performance prior to the policy stage. The question becomes: can the dissemination of information through GPIs or other IO reports inspire greater effort than the baseline?

Define $\hat{p}_\theta = P(m = 1|\theta)$ as the voter's conjecture about the probability that leader-type θ sent an ambitious report to the IO. The voter's posterior belief now incorporates two pieces of information, the IO's report s and the policy outcome x , and can be written as $\mu(x, s)$. Given voter conjectures \hat{a}_θ and \hat{p}_θ , it is optimal to retain the incumbent leader whenever:

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

where $\hat{m}_\theta = \hat{p}_\theta \phi(\sqrt{\tau}(s - 1)) + (1 - \hat{p}_\theta) \phi(\sqrt{\tau}s)$ is the total probability of observing the IO's report s given type θ .

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

the voter's reelection cutoff.

As in the benchmark model, the voter's threshold induces a probability that the incumbent will be reelected, which generates the motivation for the leader to exert effort into reforms. This reelection probability is equal to

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

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

$$\underbrace{\sqrt{\zeta}\phi(\sqrt{\zeta}(a_\theta^* - \hat{x}(a^*, \hat{p})))\Psi}_{\text{marginal electoral benefit given GPI}} = \underbrace{\lambda_\theta a_\theta^*}_{\text{marginal policy costs}}.$$

It is clear from the equation above that the IO's report affects the incumbent's selection of effort through the voter's beliefs about her type. To understand the conventional wisdom surrounding international benchmarking, it is useful to first analyze the case in which the voter's beliefs about the leader's report are held fixed. If we assume that leaders separate at the reporting stage—that aligned and misaligned leaders report their willingness to undertake governance reforms with differential probabilities—then the IO's report may successfully act as a screening device. Under these fixed beliefs, the introduction of the GPI can theoretically inspire greater domestic effort than the baseline model.

Then, to see how the IO affects effort choices, note that a strong report from the IO (strongly positive s) or a very weak report from the IO (strongly negative s) sends the voter's cutoff toward the extremes, $\hat{x}(\hat{a}, \hat{p}) \rightarrow \pm\infty$. If the voter is particularly lenient or stringent, then the incumbent leader knows the electoral outcome with near certainty, and the election is thus not conditioned on her effort choice. In these cases, the incumbent exerts no effort. However, when s takes intermediate values, downstream effort is still a valuable means of signaling type to the voter. In fact, for intermediate s there are cases in which the

leader exerts greater effort than in the model without the IO. If s is neither too high nor too low, meaning the voter gleans some information about the leader's type from the IO's report, then the leader has additional incentives to ensure that the voter perceives her as aligned, and thus exert greater effort.

Proposition 2. *Let \tilde{a}_θ^* be the equilibrium effort in the model with international reporting, and a_θ^* be the equilibrium effort without international reporting.*

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

However, this result relies entirely on the assumption that leaders' *ex ante* commitments are exogenous. By elucidating that the commitments forming the basis of the GPI are the result of strategic manipulation, we see that the conventional wisdom breaks down.

Proposition 3. *In the unique equilibrium of the model with international reporting, both leaders send $m = 1$ with probability $p^* \in (0, 1)$. Optimal effort is equivalent in both models.*

How do leaders determine which message to send to the IO? There are no direct costs to submitting ambitious ($m = 1$) or weak ($m = 0$) governance plans; the costs and benefits manifest entirely through the voter's posterior beliefs and the subsequent effect on reelection. Therefore, the leader's messaging strategy relies on maximizing her expected electoral survival over the distribution of the IO's potential reports:

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

The IO attempts to act as a screening device—separating aligned reformers from misaligned types by creating partially informative scorecards. However, leaders anticipate the

downstream effect of this information revelation. The key factor is that leaders' messages must be mutually consistent with the beliefs the voter holds downstream. To see why, consider how the IO's report affects the voter's reelection threshold. If the voter believes that aligned leaders were more likely to send optimistic messages than misaligned leaders and the IO's report is poor, then the voter sets a stringent reelection cutoff. This is not beneficial for aligned leaders. By contrast, if the voter believes that misaligned leaders are more likely than aligned leaders to report $m = 1$, then a very strong IO report engenders a stringent cutoff, which harms misaligned leaders. With an eye toward reelection, leaders seek to ensure that the voter uses the most favorable threshold to assess type, which on average occurs if the voter believes each type is equally likely to send $m = 1$.

The presence of the IO creates a chase-and-evade logic between aligned and misaligned leaders. Because talk is cheap, it is always in the interest of misaligned leaders to mimic aligned leaders by sending $m = 1$. Misaligned leaders will submit lofty proposals of reform to the IO without any intention of seeing such policies through (Melnick and Smith 2025). Conversely, the aligned leader—in an attempt to retain her *ex ante* electoral advantage—wishes to separate herself from the mimicking misaligned type. To do so, she counterintuitively prefers to underestimate her willingness to reform as a means of distinguishing herself vis-à-vis the voter. Hence, the leader's messaging is chosen purely to optimize domestic electoral attractiveness: misaligned leaders prefer to pool to conceal their bias, while aligned leaders want to separate. This dynamic echoes classical incentives to misrepresent elsewhere in international politics, where leaders may find it strategically beneficial to overstate their costs or misrepresent their type to achieve a domestic or bargaining advantage (Fearon 1995; Schultz 1998; Powell 1999).

Given these competing temptations, the only mutually consistent equilibrium strategy is for leaders to randomize between sending either message, and to do so with equal probability. Such a strategy is the only means through which both leaders can optimally satisfy their

domestic electoral constraints. Unfortunately, this strategic randomization renders the IO's report completely uninformative. The voter completely discounts the GPI in his estimation of leader type, and the equilibrium level of downstream effort reverts exactly to that of the baseline model without the IO. In short, the domestic political survival problem wholly subordinates the international institutional goal. Leaders strategically co-opt the international reporting mechanism to stifle any downstream informational effects, demonstrating that voluntary benchmarking cannot systematically inspire greater reform effort.

While the model assumes that reporting to the IO is costless and free of direct reputational consequences, the core mechanism of strategic obfuscation is robust to relaxing these assumptions. As demonstrated in the Appendix, introducing a costs for making ambitious commitments ($m = 1$) or a reputational penalty for poor IO reports (low s) merely shifts the equilibrium mixing probabilities without altering the qualitative results. The chase-and-evade dynamic persists, confirming that the domestic survival motive endogenously neutralizes the informational value of voluntary reporting.

Conclusion

The proliferation of voluntary reporting frameworks and global performance indicators has been driven by the notion that international transparency empowers domestic accountability. By providing voters with standardized metrics of a government's commitments, international organizations ostensibly reduce information asymmetries and incentivize leaders to exert greater effort toward good governance.

By formalizing the strategic interaction between international benchmarking and domestic electoral survival, this paper instead demonstrates that the introduction of a voluntary reporting mechanism does not systematically inspire greater reform effort. In fact, the very mechanism designed to enhance transparency, costless self-reporting to an IO, induces strate-

gic obfuscation. Because an ambitious international pledge signals a preference for reform to domestic audiences, misaligned leaders have a dominant incentive to mimic aligned leaders by submitting robust, yet hollow, governance plans. To preserve their electoral advantage, aligned leaders are forced to randomize their own commitments in a strategic chase-and-e evade dynamic. Ultimately, leaders endogenously neutralize the informational capacity of the IO, rendering its scorecards uninformative to the voter.

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

Crucially, the shortcomings of these benchmarking frameworks stem from their structure, not their strength. In the model, the precision of the IO's measurement capacity, τ , plays no role in shaping equilibrium outcomes. This contrasts with extant studies emphasizing variation in institutional strength as a key determinant of compliance ([Dai 2002](#); [Hafner-Burton 2008](#)). When IOs rely on self-reported data, leaders cognizant of their reelection odds will strategically obfuscate.

Consequently, future research and policy design could pivot away from building inter-

national reporting architectures on top of domestic policymaking, and instead focus on the parameters that actually drive accountability in equilibrium. IOs could use their expertise to directly facilitate domestic transparency, such as monitoring bureaucratic quality or funding independent civil society (ζ), which ensures voters have a clearer mapping of actual reform effort to policy outputs (Slough 2024). Furthermore, IOs can seek to shape the public's baseline policy preferences (λ_1) through public-facing campaigns that do not rely on information collected from national governments. In this way, international organizations can facilitate good governance not by creating new opportunities for leaders to strategically obfuscate, but by strengthening the preexisting domestic electoral incentives that actually bind them.

References

- Ashworth, Scott and Kenneth W. Shotts. 2010. “Does informative media commentary reduce politicians’ incentives to pander?” *Journal of Public Economics* 94(11):838–847.
- Awad, Emiel and Nicolás Riquelme. 2024. “International Cooperation, Information Transmission, and Delegation.” *Quarterly Journal of Political Science* 19(2):217–242.
- Battaglini, Marco and Bård Harstad. 2020. “The Political Economy of Weak Treaties.” *Journal of Political Economy* 128(2):544–590.
- Bayer, Patrick and Lorenzo Crippa. 2025. “Government Influence in Information Production in International Organizations.”
- 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.
- Bueno de Mesquita, Bruce, Alastair Smith, Randolph M. Siverson and James D. Morrow. 2003. *The Logic of Political Survival*. Cambridge: The MIT Press.
- Buisseret, Peter and Dan Bernhardt. 2018. “Reelection and Renegotiation: International Agreements in the Shadow of the Polls.” *American Political Science Review* 112(4):1016–1035.
- Chaudoin, Stephen. 2014. “Promises or Policies? An Experimental Analysis of International Agreements and Audience Reactions.” *International Organization* 68(1):235–256.
- Dai, Xinyuan. 2002. “Information Systems in Treaty Regimes.” *World Politics* 54(4):405–436.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Honig, Dan. 2018. *Navigation by Judgment: Why and When Top-Down Management of Foreign Aid Doesn’t Work*. Oxford, New York: Oxford University Press.

Honig, Dan. 2019. “When Reporting Undermines Performance: The Costs of Politically Constrained Organizational Autonomy in Foreign Aid Implementation.” *International Organization* 73(1):171–201.

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

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.

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

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

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

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

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

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

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

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

Prat, Andrea. 2005. “The Wrong Kind of Transparency.” *The American Economic Review* 95(3):862–877.

- Roberson, Brian. 2006. "The Colonel Blotto game." *Economic Theory* 29(1):1–24.
- Rosendorff, B. Peter. 2005. "Stability and Rigidity: Politics and Design of the WTO's Dispute Settlement Procedure." *The American Political Science Review* 99(3):389–400.
- Rosendorff, B. Peter and Helen V. Milner. 2001. "The Optimal Design of International Trade Institutions: Uncertainty and Escape." *International Organization* 55(4):829–857.
- Schultz, Kenneth A. 1998. "Domestic Opposition and Signaling in International Crises." *The American Political Science Review* 92(4):829–844.
- Simmons, Beth A. 2009. *Mobilizing for Human Rights: International Law in Domestic Politics*. Cambridge: Cambridge University Press.
- Slough, Tara. 2024. "Bureaucratic Quality and Electoral Accountability." *American Political Science Review* 118(4):1931–1950.
- Von Stein, Jana. 2005. "Do Treaties Constrain or Screen? Selection Bias and Treaty Compliance." *The American Political Science Review* 99(4):611–622.